

P

SEARCH REQUEST FORM
Scientific and Technical Information Center

Access DB#

51641

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: Toby Port

Searcher Phone #: 308-3531

Searcher Location: _____

Date Searcher Picked Up: 9/26

Date Completed: 9/26

Searcher Prep & Review Time: 10

Clerical Prep Time: _____

Online Time: 10

Type of Search

NA Sequence (#) 1

AA Sequence (#) 2

Structure (#) _____

Bibliographic _____

Litigation _____

Fulltext _____

Patent Family _____

Other _____

Vendors and cost where applicable

STN _____

Dialog _____

Questel/Orbit _____

Dr.Link _____

Lexis/Nexis _____

Sequence Systems CS

WWW/Internet _____

Other (specify) _____

PTO-1590 (1-2000)

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

```
US-09-615-285-1  
1738  
Title: 1 ggcgagagcgagagcgagag.....ctgcgaacaa  
Perfect score: 1 ggcgagagcgagagcgagag.....ctgcgaacaa  
Sequence: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Scoring table:  
1738 segs, 7733874588 residues  
Parameters:
```

```
Searched:
Total number of hits satisfying constraints: 0
Minimum DB seq length: 2000000000
Maximum DB seq length: 2000000000
Minimum Match 0%
Maximum Match 100%
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

Genemb1.*
1: qb_ba1.*
2: qb_ba2.*
3: qb_ba3.*
4: qb_in1.*
5: qb_in2.*
6: qb_in3.*
7: qb_cm.*
8: qb_ov.*
9: qb_pat1.*
10: qb_pat2.*
11: qb_ph.*
12: qb_pl1.*
13: qb_pl2.*
14: qb_pl3.*
15: qb_p4.*
16: em_ba2.*
17: em_ba3.*
18: em_fun.*
19: em_hum.*
20: em_hgo_inv.*
21: em_hgo_rod.*
22: em_hg_hum2.*
23: em_hg_hum3.*
24: em_hg_hum4.*
25: em_hg_hum5.*
26: em_hg_hum6.*
27: em_hg_hum7.*
28: em_hg_hum8.*
29: em_hg_inv1.*
30: em_hg_inv2.*
31: em_hg_inv3.*
32: em_hg_inv4.*
33: em_hg_inv5.*
34: em_hg_inv6.*
35: em_hg_inv7.*
36: em_hg_inv8.*
37: em_hg_inv9.*
38: em_hg_inv10.*
39: em_hg_inv11.*
40: em_hg_inv12.*
41: em_hg_inv13.*
42: em_hg_inv14.*
43: em_hg_inv15.*

44: em.ov:*
45: em.ph:*
46: em.pl:*
47: em.ro:*
48: em.sts:*
49: em.sy:*
50: em.un:*
51: em.v1:*
52: qb.sts1:*
53: qb.sts2:*
54: qb.sts3:*
55: qb.sy:*
56: qb.un:*
57: qb.v11:*
58: qb.v12:*
59: qb.hc91:*
60: qb.hc92:*
61: qb.hc93:*
62: qb.hc94:*
63: qb.hc95:*
64: qb.hc96:*
65: qb.hc97:*
66: qb.hc98:*
67: qb.hc99:*
68: qb.hc910:*
69: qb.hc911:*
70: qb.hc912:*
71: qb.hc913:*
72: qb.hc914:*
73: qb.hc915:*
74: qb.hc916:*
75: qb.hc917:*
76: qb.hc918:*
77: qb.hc919:*
78: qb.hc920:*
79: qb.hc921:*
80: qb.hc922:*
81: qb.hc923:*
82: qb.hc924:*
83: qb.hc925:*
84: qb.pr1:*
85: qb.pr2:*
86: qb.pr3:*
87: qb.pr4:*
88: qb.pr5:*
89: qb.pr6:*
90: qb.pr7:*
91: qb.pr8:*
92: qb.pr9:*
93: qb.ro1:*
94: qb.ro2:*
95: qb.ro3:*
96: qb.ro4:*
97: qb.ro5:*
98: em.bas:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the total score distribution.
and is derived by analysts

SUMMARIES

Description

no sapl

[illegible]

CDS	114..1592	/gene="TMPRS52"	/codon_start=1	/product="androgen-regulated serine protease TMPRS2 precursor"	/protein_id="AAK29280.1"	/db_xref="GI:13540004"	/translation="MALNSGPPALGPYVENHGYENPYPAPQTVPTVEYVHPADQYPSPPQYAPRYLQANPNVPCVQPKSPSGYVQTSKTKALCITLTLEFLVGLAALAGILMEFWGSKCSNGSIECDSSGTCINPSNMNCDGVSHCPGGEDENCVRLGYNFILOYVSQKRWHPVQDDMDNENYGRACRBMGTKNFYSQCIYVDSGTSFPMKLTNSAGYDIQKLYHSDACSSKAAVSLRCIACGVNLNNSRQRIYGGSSALPGAMPQVSLIHSNVHVGGSITIPPEWIVTAHCVEKPLNNPWHMTAFGILLROSFMRYGAGYOEKVISPNVDSKKNNDIALOKPLTEFNDLVKPYCLNPMPMLQBPOLCMISSMGWTEBERGTSSEVNAAKVLLITFQTRNSRYVDNLIRPAMICAEFLQGNVDSGDSGGLVTSYKSNITWLLGDTSMGSCAKAYRPRGYGNMVFMTDILYQNRADG"
	879..1589	/gene="TMPRS52"	/product="androgen-regulated serine protease TMPRS52"				
mat_peptide							
BASE COUNT	434 a 459 c 474 g 373 t						
ORIGIN							
Query Match	100.0%; Score 1738; DB 89; Length 1740;						
Best Local Similarity	100.0%; Pred. No. 0;						
Matches 1738; Conservative	0; Mismatches 0; Indels 0; Gaps 0						
QY	1	ggcggagcgagcgagcgagcgagggcgagggcgagccgccttgagcgcgagcagtcacat	60				
DB	3	ggcggagcgagcgagcgagggcgagggcgagggcgagccgccttgagcgcgagcagtcacat	62				
QY	61	attgacatccagatccatcatcattactcgaatgctgtgtgataacagcagaatggcttgg	120				
DB	63	attgacatccatccatcatcattactcgaatgctgtgtgataacagcagaatggcttgg	122				
QY	121	aactcagggtaccacacacagctatttggagccttactatgaaanaacatcggataccacaggaa	180				
DB	123	aactcagggtaccacacacagctatttggagccttactatgaaanaacatcggataccacaggaa	182				

[illegible]

Db 663 GCCTGCAGGACATGGGCTATAAGAAATATTTTACTCTAGCCAGGAATAGTGATGAC 722
Qy 721 agcgaggtccacagcgtttatgaactgacacaaagtgcggcaatctcgatctctataa 780
Db 723 AGCGGATCCACAGCTTTATGAAATGAAACACAAAGTCCGGCAATGTGATATCTATAAA 782
Qy 781 aaacttacacagtgatgcctctcttcaaaagcagtggttcttctaacctctataacc 840
Db 783 AAACCTTACCACAGTATGCTCTTTCTTCAAAAAGCAGTGGTTCTTTAGCCTGTATAGCC 842
Qy 841 tgcgggggtcaacttgaactcaaacgcgcagcagagatitgtggcgcgagagcgcgctc 900
Db 843 TCGGGGGTCAACTTGAATCAAAACCCCGCAGAGAGATGTGGCGCGGAGAGCGGCGTC 902
Qy 901 ccgggggcttgccctgcaagtgatcaactgcgaagtcgaagtcgaagtcgaagtcgaagtc 960
Db 903 CCGGGGGCTTGCGCTTGAGAGGTGACCTGCAAGCTCCAGAACTCCAGTGTGCGGAGGC 962
Qy 961 tccatcatcaaccccgagtgatcgtagacagccgcccactgcgtggaaaaactctaac 1020
Db 963 TCCATCATCAACCCCGAGTGGATCGTAGACGCCGCCACTGCGTGGAAAAACCTTTAAC 1022
Qy 1021 aatccatggcattgagcggcatttgcgggattttagaacaalcttcatlytctatgga 1080
Db 1023 AATCCATGGCATTGGACGCAATTTGCGGGGATTTGAGACAAATCTTTCATGTTCTATGGA 1082
Qy 1081 gccggataccagtagaaaaagtatttctcatccaatattagctccaaagccaagaac 1140
Db 1083 GCGGGATACCAAGTAAAGAAAGATTTCTCAACCAAAATTAATGACCAAGACCAAGAAC 1142
Qy 1141 aatgaatctgctgtagaagctgcgagagcctctgaacttcaagcagctagtgaaacca 1200
Db 1143 AATGACATTGCGCTGTATGAAAGCTGCAGAAAGCCCTCTACTTTTCAACCACTAGTAAACCA 1202
Qy 1201 gtcgtctgcgcaaaccaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1260
Db 1203 GGTGTCTCTCCCAACCCAGGATGATGCTGACGCCAGAACAGCTGTGCGGATTTCCGGG 1262
Qy 1261 tggggggcagcag 1320
Db 1263 TGGGGGGCCACCGAGGAGAAAGGAGACCTCAGAAAGTCTGAAACCTGCCAAGGTGCTT 1322
Qy 1321 ctcatttgaagacagagatgacacagcagacagatgtctatgaacaacttgatcaaccagcc 1380
Db 1323 CTCATTGAGACACAGAGATGCAACACATATGCTATGACAACTGTATCAACCAAGCC 1382
Qy 1381 atgattctgtgcggtctctcgaaggggaaagcgtgcgattcttgcaggggtgacagtgaggg 1440
Db 1383 ATGATCTGTGCGGGCTTCTGCAAGGGAGCGTCGATTTTTCGCAAGGGTGACACTGGAGGG 1442
Qy 1441 cctcttggttacttcgaagaacaatctctgtgtgctgataaggatacaagctggggctct 1500
Db 1443 CCTGTGCTACTTTCGAAAGCAATATCTGTGTGCTATGAGGATTCGAAGGTGGGTTCTT 1502
Qy 1501 ggcgtgtgcaaaagcttacagcagcagagtgatcgaggaatgtagatgtagatgtagatgtag 1560
Db 1503 GCGGTGCGCAAAAGCTTACAGACCAAGAGGTGACGGGAATGTGATTTTACAGGACTGG 1562
Qy 1561 attatcgaacaaatgagggcagagcaggttaatacgaatggtcttgcgtccttgacgtctt 1620
Db 1563 ATTATTCGACAAATGAGGCAAGCGCTAATCCACATGCTTCGTCCTTGAAGCTGTTT 1622
Qy 1621 tacaagaacaaatgagggctgt 1680
Db 1623 TACAAGAAAACAAATGGGCTGT 1682
Qy 1681 tcaagagtgactcatcttctatcaaacagtgaaactgtctgcaaaaaaataaaaaa 1738
Db 1683 TCAGAGGTCACTTCAATTTTATTAACACATGAACTGTGTGCAAAAAAATAAAAAA 1740

RESULT 2
HSU75329

LOCUS HSU75329 2479 bp mRNA PRI 10-OCT-1997
DEFINITION Human serine protease mRNA, complete cds.
ACCESSION U75329
VERSION 075329.1 GI:2507612
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE Paoloni-Giacobino, A., Chen, H., Peltisch, M.C., Rossier, C. and Antonarakis, S.E.
JOURNAL Cloning of the TMRP52 gene, which encodes a novel serine protease with transmembrane, ID1RA, and SRCH domains and maps to 21q22.3
MEDLINE Genomics 44 (3), 309-320 (1997)
REFERENCE
AUTHORS Paoloni-Giacobino, A., Chen, H. and Antonarakis, S.E.
JOURNAL Direct Submission
TITLE Submitted (17-OCT-1996) Medical Genetics, University of Geneva Medical School, 1 Michel-Servet Street, Geneva 1211, Switzerland
FEATURES
source
1..2479
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
57..1535
/product="serine protease"
/protein_id="AAC51784.1"
/db_xref="GI:2507613"
CDS
578 a 650 c 677 g 574 t
ORIGIN
Query Match 95.2%; Score 1655.2; DB 97; Length 2479;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 56 gataatgaacattccagatactcatctactgagtgtgtgtgtaacagaaatg 115
Db 1 gtcattttgaacatttcagatattctatctgattgctgtgtgtaacagaaatg 60
Qy 116 ctttgaactcaaggtgacacacagcatttggaccttactatgaacaacatgataccac 175
Db 61 ctttgaactcaggggtacacacacagcatttggaccttactatgaacaacatgataccac 120
Qy 176 cggaaaaaccctatcccgacacagcccaactgtgttcccacgtctcagagtgatcccg 235
Db 121 CGGAAAACCCCTATCCCGACACAGCAGTGTGTCCTCCACTGCTACGAGGTGATCCGG 180
Qy 236 cttaagtaactaccggtcccggtgacccagtagcccgagggttctcgaagcaggtctca 295
Db 181 cttaagtaactaccggtcccggtgacccagtagcccgagggttctcgaagcaggtctca 240
Qy 296 acccgcgtctgacgagcccaaatcccatccggagcaggtgtgacactcaaaacta 355
Db 241 ACCCGCTCGTTCAGAGCAAGCCCAATCCCATCCGGAGAGTGTCTACTCAAAACTA 300
Qy 356 agaaagcactgtgacatcacttgaacccttgaggaccttctcgttggagctgcgtgcg 415
Db 301 AGAAAGCACTGTGCATCACCTTGACCTGGGAGCTTCTCGTGGAGAGCTGGCGCG 360

```

OY 416 ctggcctactctggaagtcaatggagcaagtgctccaactctggagtagagtgagct 475
    |||||||
DB 361 ctggcctactctggaagtcaatggagcaagtgctccaactctggagtagagtgagct 420
OY 476 cctcagatcctgcaaacccctcctaacttgtagagggctgtagaacctccccggcg 535
    |||||||
DB 421 cctcagatcctgcaaacccctcctaacttgtagagggctgtagaacctccccggcg 480
OY 536 gggagagacgaagatcggtgtgtctgcctctacgagcaaacctcctcctcagtgta 595
    |||||||
DB 481 gggagagacgaagatcggtgtgtctgcctctacgagcaaacctcctcctcagtgta 540
OY 596 catctcagaagaatctcctgtagcctcctgtgtgcagaagcctcggaggaactacggcg 655
    |||||||
DB 541 catctcagaagaatctcctgtagcctcctgtgtgcagaagcctcggaggaactacggcg 600
OY 656 gggcgagcctcggaggaacatgggtatagaatatttctaacttagcgaaggaatagtg 715
    |||||||
DB 601 gggcgagcctcggaggaacatgggtatagaatatttctaacttagcgaaggaatagtg 660
OY 716 atgacagcgaatcccaacagcttcatgaaactgaaacaagtcggcgcaatgtcagatct 775
    |||||||
DB 661 atgacagcgaatcccaacagcttcatgaaactgaaacaagtcggcgcaatgtcagatct 720
OY 776 atcaaaactctgacacagtgatgctgtctctcaaaagcagtggtctcttcaagctgta 835
    |||||||
DB 721 atcaaaactctgacacagtgatgctgtctctcaaaagcagtggtctcttcaagctgta 780
OY 836 tagcctcggaggtctcaactgaaactcaagcgcgacagagagatgttgggcgcgagagcg 895
    |||||||
DB 781 tagcctcggaggtctcaactgaaactcaagcgcgacagagagatgttgggcgcgagagcg 840
OY 896 cgcctccggggggcctgagcctctgagagtgacgtcgcagctccagaacgtgccagtg 955
    |||||||
DB 841 cgcctccggggggcctgagcctctgagagtgacgtcgcagctccagaacgtgccagtg 900
OY 956 gaggctcccatctcaaccccgagtgtagatcgtagacagcggcccaactgctgtagaaacccc 1015
    |||||||
DB 901 gaggctcccatctcaaccccgagtgtagatcgtagacagcggcccaactgctgtagaaacccc 960
OY 1016 tcaacaatccatgagcatgtagacggcacttgcgggagcttctgagacaactcttcatgtct 1075
    |||||||
DB 961 tcaacaatccatgagcatgtagacggcacttgcgggagcttctgagacaactcttcatgtct 1020
OY 1076 atggagcggagatccaaagtagaagtagatctctcatccaaatctgaactcccaagca 1135
    |||||||
DB 1021 atggagcggagatccaaagtagaagtagatctctcatccaaatctgaactcccaagca 1080
OY 1136 agaaacaatgacatctgagcctgtagagagctgtagagagcctctgacttcaagagcctag 1195
    |||||||
DB 1081 agaaacaatgacatctgagcctgtagagagctgtagagagcctctgacttcaagagcctag 1140
OY 1196 aaccagtgctctgacccaacccagagcatgtagctgtagcagcaagaacgtcctgtagatct 1255
    |||||||
DB 1141 aaccagtgctctgacccaacccagagcatgtagctgtagcagcaagaacgtcctgtagatct 1200
OY 1256 ccggggtggggggccacagagaggaaggaagcctcagaagtgtctgaaacgtctgcaag 1315
    |||||||
DB 1201 ccggggtggggggccacagagaggaaggaagcctcagaagtgtctgaaacgtctgcaag 1260
OY 1316 tgcctctcaatctgagacacagagatgtagaagcagatagtctcatgacaacctgtagaac 1375
    |||||||
DB 1261 tgcctctcaatctgagacacagagatgtagaagcagatagtctcatgacaacctgtagaac 1320
OY 1376 cagcaatgtagctgtagcggcctcctcctcaggggaagctgtagatctctcaggggtgagcag 1435
    |||||||
DB 1321 cagcaatgtagctgtagcggcctcctcctcaggggaagctgtagatctctcaggggtgagcag 1380
OY 1436 gaggcctctgtagcctcctcagaaacaatctgtagcagtagaaggggaataagcctgagg 1495
    |||||||
DB 1381 gaggcctctgtagcctcctcagaaacaatctgtagcagtagaaggggaataagcctgagg 1440
OY 1496 gttcgtggtgctgcaaaagcttcaagacacaggaagttaggggaatgtgtagatctcaagg 1555

```

```

DB 1441 gttctggctgtcccaaaagcttaccagaccagagtgtagcgaatgtgaggtattccacgg 1500
    |||||||
OY 1556 actggaattatcgacaatgagagcagagcttaatccacatggtctctcctcagat 1615
    |||||||
DB 1501 actggaattatcgacaatgagagcagagcttaatccacatggtctctcctcagat 1560
OY 1616 cgtttacagaagaacaatgagggcgtgtgtgtctcccggtgcatgattactcagaag 1675
    |||||||
DB 1561 cgttttacagaagaacaatgagggcgtgtgtgtgtctcccggtgcatgattactcagaag 1620
OY 1676 atgattcagaagcttaacttaatttataaacaagtgaaactgtctgcgc 1723
    |||||||
DB 1621 atgattcagaagcttaacttaatttataaacaagtgaaactgtctgcgc 1668

RESULT 3
AX041973
LOCUS AX041973 3966 bp DNA PAT 23-NOV-2000
DEFINITION Sequence 3 from Patent WO0065067.
ACCESSION AX041973
VERSION AX041973.1 GI:11340738
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3966)
AUTHORS Nelson,P.S., Hood,L. and Lin,B.
TITLE Prostate-specific polynucleotides, polypeptides and their methods
of use
JOURNAL Patent: WO 0065067-A 3 02-NOV-2000;
The University of Washington (US)
FEATURES
source
CDS
1..1535
57..1535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17064.1"
/db_xref="GI:11340738"
/translation="MALNSGSPALGRIYEHNGYOPENPYPAGPVPVPTVEVPAQY
TPSPVQAPVPLVTOASNPVYCTOPKSPSGTVCTSKTKALCLITLTLTVLVAALAA
GLMKFMGSKNSGIECDSSGCTINPNCMDGVSHCGEDENRNVLYGPNFLDM
YSSORFQSWHPKCODMNNENYGRACRDGYNKNNYSOGIYDDSGSTFPMKLNISAGN
VDIYKLIYSPACSSKAVSLRGLACGVNLNSRSLRVGGESALPGAMPVQVSLHYO
NVHCGGSLTPPEWIVTAHCVKPELNNPMLHTAPAGILRQSPFEGAGYQVORVISH
PNYSKTYNNIDIALMKLOKPLTENDLYAPVCLPNPAGMLQPEOLCHISMGATIEKKG
TSEVLNAKVLILIEFORCNRYVDNLITPAMICAGFLQGVNDSCQDSDGSGPLVTSNN
NIMWLIQDTSMGSGCAKAYRPGVGNVVFYDWIYROMKANG"
BASE COUNT 992 a 998 c 988 g 984 t 4 others
ORIGIN
Query Match 95.2%; Score 1655.2; DB 9; Length 3966;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
OY 56 gtcataatgacaatcccaatactatactatctcgtgtagtgctgtagaagcagaagatgg 115
    |||||||
DB 1 gtcataatgacaatcccaatactatactatctcgtgtagtgctgtagaagcagaagatgg 60
OY 116 ccttgaaactcaggggtacacacagcattatgagccttactatgaaacaatgtagataccaac 175
    |||||||
DB 61 ctttgaaactcaggggtacacacagcattatgagccttactatgaaacaatgtagataccaac 120
OY 176 cggaaaaaccctatacccgacagcccaactgtgtccccaactgtctcagaggtgtcaltccgg 235
    |||||||
DB 121 cggaaaaaccctatacccgacagcccaactgtgtccccaactgtctcagaggtgtcaltccgg 180
OY 236 ctccagtaaccctgtcccccgtgccccagtagcccccgagggttctctgacgagagcttcca 295
    |||||||

```

us-09-615-285-1.rge

[illegible]

Db	961	TTCTATGAGCCGGATTACCAAGTAGAAAAAGTGAATTTCATCAACAAATTATGACTCCAAAG	1020
OY	1132	accagaacaatgatcatctgctgataagctgcagaagcctctgactcttcaacagacctta	1191
Db	1021	ACCAAGAACAAATGACATTGGCTGATGTAAGCTGGAGAAGCCCTCTGACTTTCACAGACCTTA	1080
OY	1192	gtgaaacccgctgtgtctgcccacaccagaagcatgtctgtcagaccagaacagactctgtcg	1251
Db	1081	GTGAACCAATGCTGTGTGCGCAACCCAGGCATGATGCTGCAGCCAGAAACAGCTCTGCTGG	1140
OY	1252	atttcgcggtgtggtgggcaccccgaggagaaagggaaagacctcaagaatgtctgaacgcctggc	1311
Db	1141	ATTTTCGGGGTGGGGGCCACCGAGGAGAAAGGAGACCTTGAGAAAGTCTGAAGCGCTGGC	1200
OY	1312	aaggctctctctttaagaacagagatatgcacagcagatatgtctatgaacactgtatc	1371
Db	1201	AAGTGTCTTCTCATGTGACACACAGATCCACACACAGATATGTATGACAAACCTGATC	1260
OY	1372	acacagaacacatgatctgtgtgcggtctctctgcaggggaacgctgatctcttgcgaaggtgac	1431
Db	1261	ACACAGCAATGATCTGTGCGGCTTCTGACAGGGGAACGTGTGATTTCTTCCAGGGGTGAC	1320
OY	1432	agctgagggcctctgtgtcactctgaagaacaatatctgtgtgtctgatatgaggaataaacg	1491
Db	1321	AGTGAGGGGCTCTGTGCTCACTTGCAGAAACAAATCTGTGTGCTGATAGGGGATACAAAGC	1380
OY	1492	tggggctctgtgtctgtgcnaagctctacagaccagaggtgtacgggaatgtgaatgtgtatcc	1551
Db	1381	TGGGCTTCTGGCTGTGCCCMAAGCTTACAGACACAGGAGTGTACGGGAATGTGATGCTATTC	1440
OY	1552	acgagactgatttatcgcacaatatgaggagcagaagcttaa	1590
Db	1441	ACGGACTGGATTATTCGCAAAATGAGGCGACAGCGCTAA	1479

LOCUS	AF199362	1755 bp	mRNA	ROD	15-FEB-2001
DEFINITION	Mus musculus	serine protease	TMPRSS2 (TMprs2)	mRNA, complete cds.	
ACCESSION	AF199362				
VERSION	AF199362.1	GI:9739018			
KEYWORDS					
SOURCE	house mouse,				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 1755)				
JOURNAL	Vaara, M.H., Porvari, K.S., Kellokumpu, S., Kyllonen, A.P. and				
PUBMED	Vihko, P.T.				
REFERENCE	Expression of transmembrane serine protease TMPRSS2 in mouse and				
AUTHORS	human tissues				
TITLE	J. Pathol. 193 (1), 134-140 (2001)				
JOURNAL	11169526				
PUBMED	2 (bases 1 to 1755)				
REFERENCE	Vaara, M.H., Porvari, K.S. and Vihko, P.T.				
AUTHORS	Direct Submission				
TITLE	Submitted (27-OCT-1999) WHO CCR, Biocenter Oulu, P.O. Box 5000,				
JOURNAL	Oulu FIN-90014, Finland				
FEATURES	location/Qualifiers				
source	1..1755				
gene	/organism="Mus musculus"				
CDS	/strain="BaBb/C"				
1	/db_xref="taxon:10090"				
	1..1755				
	/gene="Tmprs2"				
	112..1584				
	/gene="Tmprs2"				
	/note="TMPRSS2 homolog"				
	/codon_start=1				
	/product="serine protease TMPRSS2"				
	/protein_id="AAF97867.1"				
	/db_xref="GI:9739018"				
	/translation="MALNSSSPGICPCYENHGYSCEHICPPRPVAFNGYNLYPAQY				

[illegible]

REFERENCE

ds: 13596 .
 File: 13596-1 .
 File: 13596-1 .
 house mouse
 Mus musculus
 Mus musculus
 Eukaryota: Eutheria
 Mammalia
 1 (bases 1 to 1153)

Accession	Gene	Species	Accession	Gene	Species
QY	1687	gcccctttcactt	QY	1687	gcccctttcactt
Db	1697	gcccctttcactt	Db	1697	gcccctttcactt

DEFINITION	Cds.	GI:6648959
ACCESSION	AF113596.1	
VERSIONS		
RECORDS	house mouse	
REPLACE	Mus musculus	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Eutheria; Rodentia	
	Mammalia	
	1 (bases 1 to 1153)	

ds: 13596 .
 File: 13596-1 .
 File: 13596-1 .
 house mouse
 Mus musculus
 Mus musculus
 Eukaryota: Eutheria
 Mammalia
 1 (bases 1 to 1153)

RESULT		
AF243500	AF243500	1356 bp transmembrane
LOCUS	Mus musculus	plasma
DEFINITION	Mus m.5500	
ACCESSION	AF243500.1	GI:7578918
KEYWORDS	house mouse,	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Eutheria; Mammalia; Chirodata; Euteleostomi; Mus musculus; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus	

REFERENCE
AUTHORS Han, J., and Trammemund
TITLE Putatively
JOURNAL unpublished
2 (basses 1 to 1735)
and Kim, S.
Han, D. and Kim, S.
REFERENCE Direct Submits (MAR-2000) Immunology, The Scripps Research Institute,
AUTHORS Submitted (USFwPines Rd., La Jolla,
TITLE Torrijon/Qualifiers
10550 N.
Local 1735
musculus"

```

FEATURES
source
1. organism="Mus musculus"
/organism="taxon:10090"
/db xref="taxon:10090"
base="

```

CDS
 /note="putref-1
 /codon_start=1
 /protein_id="AA66186.1"
 /protein_1gi="7578919"
 /db_xref="1:7578919"
 /translat="MALNSGSPGPGIGCYENHGYSEHSLICLPAPPVAPNGNYLPAQY
 LPFVWQSTSEMEGSGSGACAKDMGKYNKROSHLIVGLNAPGGMWQVLSH
 YAMFQNSPCYCODMSERLCEICGVRSKRAFLNGLDDEMSIGWATLEKSK
 SOKRANIPSCSSRWVVAEYEPILSGRPYKAFAPLILGYSGRHROVEKYL
 DYKLLHPTEWYIYLAHCLEADFLKQVCLPANGMLDILGYSVCSGGGGLYLLK
 HVGGSILNIDIALMLKQIPALNLTITPMALGYOUMRANS"
 YSKTKNNDILPEYKSCNSKYTGNTVFTDMLG
 DYNLAAMSGSCAKALPGYGVNTVFTDMLG
 457 9 398 L
 MALIGDLC

BASE COUNT ORIGIN	DB 94; Score 969; Dev. No. 3.3e-247; 380;	Length; Indels	22; Gaps
55.8%;			151

[illegible][illegible]

Db	1455	GGCGATTGGGGACACAGCTGGGGCTCGGCGTGTGCCAAGGCACCTCAGACCTGGAGTAT	1514
Qy	1532	acggaagatgtagtatacgaagactggaattatcgacaaatgaagagcgaacgagtaat	1591
Db	1515	ACGGGAACGTACGTCGTTATTAACAAGTTGAGTATACACGCAGAAATGAGCGCAACAGCTAAT	1574
Qy	1592	ccacaatg-----tcctgcaccttgagtcgcttttcaagaanaacaatg	1636
Db	1575	CCACGTGGCTTGTGCCAGACTTCCTTTGCTTTCACAAACCTTCGCAGAGAAAACCAAG	1634
Qy	1637	gagctg-gtttcgctcccccgtgacatgaattacatctagaagatcatcaagagcatcca	1695
Db	1635	GCGCTAATTTTAAGTCTCTGTCGACAAATGTACCTTTTGAGATGATTCGAGGCGCTTTCA	1694
Qy	1656	ctttattaaacagatgaactgctctggcaaaaaaaaaa	1736
Db	1695	CTTTTACTTAACAGTGAAGCTTGTTCGACTGTGAAAAA	1735

	RESULT	8	
	LOCUS	AB038497	
	DEFINITION	Xenopus laevis xesp-2 mRNA for embryonic serine protease-2, complete cds.	VRT 09-AUG-2000
	ACCESSION	AB038497	
	KWORDS	embryonic serine protease-2; Xenopus embryonic serine protease-2	
	SOURCE	Xenopus laevis	
	ORGANISM	Xenopus laevis	

REFERENCE AUTHORS TITLE	JOURNAL MEDLINE REFERENCE
1 (sites) Yamada, K., Takabatake, T. and Takeshima, K. Isolation and characterization of three novel serine protease genes from <i>Xenopus laevis</i> Gene 252 (1-2), 209-216 (2000)	20363741
2 (bases 1 to 3507) Yamada, K., Takeshima, K. and Takabatake, T. Direct Submission Submitted (17-FEB-2000) to the DDBJ/EMBL/GenBank databases. Kazuo Yamada, Nagoya University, Graduate School of Human Informatics, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan (E-mail: yamada@atleinfo.human.nagoya-u.ac.jp, Tel.: +81-52-789-2573)	

FEATURES	source	Location/Qualifiers
gene	1. .3507	/organism="Xenopus laevis"
gene	/db_xref="taxon:8355"	
gene	512. .2815	
gene	/gene="Xesp-2"	
gene	512. .2815	
gene	/gene="Xesp-2"	
gene	/note="LDLR domain	
gene	/note="protease"	
gene	/codon_start=1	
gene	/product="embryonic serine protease-2"	
gene	/protein_id="BAB08217.1"	
gene	/db_xref="GI:9757700"	
gene	/translation="MEPAVYIYPDNNELEQPRVYDVYNSMTQVAPALPYNTQVSHSNQV MASPISEFQMSFRKRVVLISGLSTIVAAVIVASVLITGFVSTTSVYKQCKGSSSK CIYGYOICDGVDCCPYGDDERNCAVTCILYCSYSTCIYVOICGVODCCPYGDDERC ATPEPTPCOMCSYTYTCIYGYOICGVODCCPYGDDERNCAVTPICOMCSY TSTCIYGYOICGVODCAAGDDERNCAVTPICOLYCSYTYTCIYAOICGVOD CEPVDERNCAVATSTPCOICYCDMPEDYTCLIAYQCMCGVRCYGYDDERNCAV TTTTATTSTPPTCOICYCMENMYTYTCIYAOCMCGVRCYGYDDELNCDFTTTAYEE KRGGSVCVLSQMDGCVDPCEGDEMCSVLSLPADFOLOYSTVSAMLPKCSGY WMDNFRFACODGETNGSSYIKRIIDTLMSRYPAKNGCFKLYSGVMSKRYTSSVYCY SCVMSVHLICISGCVSNMSLVSRIVGGTFRANIGAMPQVONLOYITTCGLGGSITSKKIL VYVAHCVGSYSASAGMFWAGCTLRKPSYVNASAFVEKRIIVHFGKSYTIDNDIALM KIDLETTFGTTQAPCLPNSMGFWAGTTTTCISGVSIEGGSVSYLYQVAITPLDSS KNCDESYTNGOITSTMCAGVLTSGGVDTCCDSCGSPVLRNKGITWMLVQITSGWDCG	

```

      .      ARANKPGVGNVTFFLEMIYSQMRITR
BASE COUNT  1035 a      712 c      740 g      1020 t
ORIGIN

```

Query Match	17.0%;	Score 295.6;	DB 8;	length 3507;
Best Local Similarity	55.2%;	Pred. No. 8.3e-68;		
Matches 621; Conservative	0;	Mismatches 499;	Indels 6;	Gaps 2;

[illegible]


```

/note="LIMB3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(14330..14682)
repeat_region
/note="MLT1A1"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
14685..15070
/note="LIMB3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(17893..18027)
repeat_region
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
18438..19329
repeat_region
/note="MER49"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
complement(20328..20688)
repeat_region
/note="MT1A1"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(20741..20997)
repeat_region
/note="LTR16C"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
21744..22035
repeat_region
/note="Alu5g"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(22282..22412)
repeat_region
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(22950..23031)
repeat_region
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(25177..25479)
repeat_region
/note="Alu5g"
/rpt_family="SINE/Alu"

```

Query Match 12.8%; Score 223; DB 92; Length 340000;
 Best Local Similarity 100.0%; Pred. No. 1.8e-48;
 Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 gggcaccacacagctatggaccttactatgaaccatgataccacggaaccc 186
 |||||
 DB 303106 GGGTCACACACACTATGGACCTTACTATGAACCATGATACCAACCGAACC 303047

QY 187 tatccggacacagccactggtgctccctgcttactagaggtgtatccggtcagtactac 246
 |||||
 DB 303046 TATCCCGCAGCCACTGTGCTCCACTGTCTACGAGGTGATCCGGCTCAGTACTAC 302987

QY 247 ccgtcccccgtgcccagtaagcccgaggtgctctagcagcaggtcccaaccccgctc 306
 |||||
 DB 302986 CGGTCCCGGTGCGCCAGTACGCCCGAGGCTCTTACGACGAGCTTCCACCCGCTGCTC 302927

QY 307 tgcagcagcccaaatcccatccggagcaggtgtgacctcaa 349
 |||||
 DB 302926 TGCACGACGACCCCAATCCCATCCGCGAGACAGTGTGACCTCAA 302884

RESULT 12
 AB038159 AB038159 2135 bp mRNA PRI 17-JAN-2001
 DEFINITION Homo sapiens TMPRSS3c mRNA for serine protease, complete cds.
 ACCESSION AB038159
 VERSION AB038159.1 GI:12246827
 KEYWORDS serine protease.
 SOURCE Homo sapiens liver and thymus cDNA to mRNA.
 ORGANISM Homo sapiens

```

REFERENCE
AUTHORS
1 (sites)
Scott,H.S., Kudoh,J., Wattenhofer,M., Shibuya,K., Berry,A.,
Chrast,R., Guipponi,M., Wang,J., Kawasaki,K., Asakawa,S.,
Machida,S., Yonush,F., Mehdi,S.O., Radhakrishna,U.,
Papavasvas,M.P., Gehrig,C., Rossier,C., Korostilovsky,M., Gal,A.,
Shimizu,N., Bonne-Famir,B. and Antonarakis,S.E.
Insertion of beta-satellite repeats identifies a transmembrane
protease causing both congenital and childhood onset autosomal
recessive deafness
Nat. Genet. 27 (1), 59-63 (2001)
JOURNAL
MEDLINE
20578749
REFERENCE
AUTHORS
Shimizu,N., Kudoh,J. and Shibuya,K.
Direct Submission
Submitted (10-FEB-2000) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp,
tel:81-3-3351-2370, fax:81-3-3351-2370)
FEATURES
Source
1..2135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/issue_type="Liver and thymus"
250..1233
/gene="TMPRSS3c"
250..1233
/gene="TMPRSS3c"
/note="TMPRSS3c"
/codon_start=1
/product="serine protease"
/protein_id="BAB20079.1"
/db_xref="GI:12246828"
/translation="MCSDDMKGHYANVACAOIGFPYVSSDNLRVSLRGPREPVS
IDHLLPDDKVTALHSHVYRBCGASHVYTCIDGHRGYSRRVGGMSLLSQMP
WQNSLDFGCHLGGSVITPMTTAAHCVYDYLKPSFTIOGLVSLDNPAPSHV
EKIVHSKTPKRLGNDIALMKLAGPLTENNIOPLVLSNSENFPDGVCWPTSGWA
TEBGADGADSPVNLHAAPLISNICKHNRVYGGIISPMICAGYLTGVDVSOOGDSGG
PLVCOERRMLKLVGATFSGICAEVVKPVTYRTVTFILMIHEOMERDLKT"
BASE COUNT 487 a 582 c 562 g 504 t
ORIGIN
```

Query Match 12.2%; Score 212.2; DB 85; Length 2135;
 Best Local Similarity 54.2%; Pred. No. 1.4e-45;
 Matches 624; Conservative 0; Mismatches 488; Indels 39; Gaps 8;

QY 469 tgcgacctcccaagttactatcatcaacccttaactggtgtatggtgtctaacctgc 528
 |||||
 DB 103 TGTGCTCATCTTTAAGTATATGACGTAGCTGATGATGACGAGTCTGGATTGC 162

QY 529 cccgagcggaggaagcgaatcggtgtgttcgccttaagcgaacactcatccttcag 588
 |||||
 DB 163 AAGAGCGGGAGGAGGACAGATGCCCTGTGCTGGGTGGTGTAGAAATGCCGTCTCAG 222

QY 589 gtgtactcatctcagaagaaagctctgagaccctgtgtgccaagaagcagcagaaac 648
 |||||
 DB 223 GTGTTTCACACTCTCT-----TCGTGGAAGACCATGATGCTCCGATGCTGAAGGGTCA 276

QY 649 taagggcgccgctgcagcaggaatggtgtataagaataattttacttaagcaagga 708
 |||||
 DB 277 TACGCAAAATGTCTCTGTGCCCACTGGGTTTCCCAAGCTATGTGAGTTAGATAACCTC 336

QY 709 atagtgtgaggaagcagcaccagcttatgaactgaacaaagtcgagcaatgt 767
 |||||
 DB 337 AGAGTAGCTCGCTGAGAGGGGAGATTCCGGAGAGAGATTGTGCATGATACACTCTTG 396

QY 768 cgatatctataaaaaactgtaccac-----agtgatgctgttcttca 810
 |||||
 DB 397 CCAGATGACAAAGTGACTGATACACACTCATATATGTAGAGGAGGAGATGTGCTCT 456

Db	790	GGCAGCGTGGTTACCTTGACGTGCACAGCGCTGTGGTCATAGAAAGGGCTACAGCTCAC--	847
Qy	871	agcagcattgttggcgcgagagcgcgtcccggyggccctgtgcagtcagtcgcctg	930
Db	848	----GCATCGCTGGGTGGAAACATGTCTTGCTCTGCAGATGGCCCTGGCAAGCCAGCTT	903
Qy	931	cagctccagaagctccagctgtgtgagatgcttcatactatcaccccccgaatgtatcgtgaca	990
Db	904	CAGTTCACAGGGCTACCAACCTGTGCGGGGGCTGTGATACAGCCCTCTGTGATATCACT	963
Qy	991	gcgcgccaccgcgttggaaaaaccttcaacatccatgcatgttcagtcagcgcatttggcggg	1050
Db	964	GTCTGCACACGTGTTTATGA--CTTGTACTCTCCCAAGTCATGGA---CAATCCAGGTGG	1018
Qy	1051	attttgacacatcttcatatgltctatgttagcgcgatatcaagtaagaaatgattctt	1110
Db	1019	GTCTAGATTTCCCTGTTGGCAATCCA--GCCCATCCCACTTGTTGGTAAGAAATTTGTCTAC	1077
Qy	1111	catecaaatattgacctccagaaccacaagaacatgatcatgtgcgtgtgaatctgcagaag	1170
Db	1078	CACAGCAGATACAGCCCAAGAGGCTGGGCATGACATGCGCTTATTAACCTGGCCGGG	1137
Qy	1171	ccttctgacttcaacagcactagtfgaaaccagtgltgtctgtcccaaccagcagatgctgt	1230
Db	1138	CCACTACAGTTCAATGAATGATCCAGCGCTGTGTGCTGTGCCAATCTGAGAGAGACTTC	1197
Qy	1231	cagccagaacagctctgtctgtgatttccgggtgtgggggtccaccgaga--gaaagggaag	1287
Db	1198	CCCGATGGAAGATGTGCTGGAGCTGAGATGGGGGGCCACAGAGATGAGCAGGTGAC	1257
Qy	1288	accctcagaagtgtctgaagcgtgcgaaggtgtcttcatattgagacacagatgtgaacagc	1347
Db	1258	GCTTCCCTTCTTGAAACCAAGCGCGGCTGCTTTGATTTCCACACAAATCTGCACACAC	1317
Qy	1348	agatatgtcctatjacaaactgtatcacaccagcagatctgtgcggtcttctcagggg	1407
Db	1318	AGGGAGCTGATAGGTGGCATCATCTCCCTCTCAGTGTCTGGGGGCTTACTGACGGGT	1377
Qy	1408	aagctgattcttgcacaggtgtgacagtgtgagggccttgytcaacttcgaagaacatatc	1467
Db	1378	GGCGTGGACAGCTGCACAGGGGGACACGGGGGGCCCTGTGTGTGCAAGAGAGAGCTG	1437
Qy	1468	tgtgtgagctgtatagggatatacaactctgggttctgtccttgcacaaagcttacaagccaga	1527
Db	1438	TGGAAATTAGTGGAGCGACGACCACTTTGGCATCGGCTGGCGAGAGGTAAACACCTGGG	1497
Qy	1528	gtgtcagggaaatgtgattgtatcaccagcagtcagtcatttcaacaaatgaggcagcgc	1587
Db	1498	GTTGATACCCGTGTACACTCTCTCTGAGCTGATGCCAGGACAGATGAGAGAGACCTA	1557
Qy	1588	taatccacatgt	1598
Db	1558	AAAACTGGAAG	1568

RESULT 14			
LOCUS	AB038158		
DEFINITION	AB038158	2554 bp	mRNA
ACCESSION	AB038158		PII
VERSION	AB038158.1	GI:12646825	17-JAN-2001
KEYWORDS	serine protease, complete cds		
SOURCE	Homo sapiens liver and thymus cDNA to mRNA.		
ORGANISM	Homo sapiens		

REFERENCE	AUTHORS	TITLE
(sites)		
Scott H.S.	Kudoh, J., Matthenofer, M., Shibuya, K., Berry A., Chirast, R., Gulipponi, M., Wang, J., Kawasaki, K., Asakawa, S., Minoshima, S., Yonnu, F., Meidli, S. O., Radhakrishna, U., Papanicolaou, S., Johann, C., Rossler, C., Korostiltshevsky, M., Gal, A., Shimizu, N., Bonne-Tamir, B. and Antonarakis, S.E.	Insertion of beta-satellite repeats identifies a transmembrane

protease causing both congenital and childhood onset autosomal recessive deafness
Nat. Genet. 27 (1), 59-63 (2001)
20578749
2 (bases 1 to 2554)
Shimizu, N., Kudo, J. and Shibuya, K.
Direct Submission
Submitted (10-FEB-2000) Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-Ku, Tokyo 160-8582, Japan (E-mail: nshimizu@med.keio.ac.jp, Tel: 81-3-3351-2370, Fax: 81-3-3351-2370)
Location/Qualifiers
I. .2554

gene
CDS

/note="transmembrane protease, serine 3"

BASE COUNT	590 a	679 c	662 g	623 t
------------	-------	-------	-------	-------

Query Match	12.28;	Score 212.2;	DB 85;	Length 2554;
-------------	--------	--------------	--------	--------------

Matches 624; Conservative 0; Mismatches 488; Indels 39; Gaps 8

QY 469 tgcgactcctcaggtacctgcatcaacccctctaactgtgtgatgycgtgtcacactgc 528

Db 522 TGTCGCTCATCCTTTAAGTGTATCGAGCTGATAGCTCGATGTGACGGAGTCTCGGATTGC 581

QY 529 cccggcgggagacgaatcgtgtgttcgcctctacgaccaacttcctcag 588

Db 582 AAAGACGGGAGGACGAGTACCGCTGTGTCCGGTGGGTGGTCAGAAATGCCGTGCTCCAG 641

589 gtgtactcatctcagagaagtccttgcaccctgtgtgccaagcgaacttgaacgagaac 648

DD 042 GTGTCACAGCTGCT-----TCGTGGAAGCATTGTGCTCCGATGACTGGAAGGTCAC 095

049 cacgugcugcccgacaggaalaacllcacllcagccaaugga / 08

cy

[illegible][illegible][illegible]

0266 0267 0268 0269 0270 0271 0272 0273 0274 0275 0276 0277 0278 0279 0280 0281 0282 0283 0284 0285 0286 0287 0288 0289 0290 0291 0292 0293 0294 0295 0296 0297 0298 0299 0300 0301 0302 0303 0304 0305 0306 0307 0308 0309 0310 0311 0312 0313 0314 0315 0316 0317 0318 0319 0320 0321 0322 0323 0324 0325 0326 0327 0328 0329 0330 0331 0332 0333 0334 0335 0336 0337 0338 0339 0340 0341 0342 0343 0344 0345 0346 0347 0348 0349 0350 0351 0352 0353 0354 0355 0356 0357 0358 0359 0360 0361 0362 0363 0364 0365 0366 0367 0368 0369 0370 0371 0372 0373 0374 0375 0376 0377 0378 0379 0380 0381 0382 0383 0384 0385 0386 0387 0388 0389 0390 0391 0392 0393 0394 0395 0396 0397 0398 0399 0400 0401 0402 0403 0404 0405 0406 0407 0408 0409 0410 0411 0412 0413 0414 0415 0416 0417 0418 0419 0420 0421 0422 0423 0424 0425 0426 0427 0428 0429 0430 0431 0432 0433 0434 0435 0436 0437 0438 0439 0440 0441 0442 0443 0444 0445 0446 0447 0448 0449 0450 0451 0452 0453 0454 0455 0456 0457 0458 0459 0460 0461 0462 0463 0464 0465 0466 0467 0468 0469 0470 0471 0472 0473 0474 0475 0476 0477 0478 0479 0480 0481 0482 0483 0484 0485 0486 0487 0488 0489 0490 0491 0492 0493 0494 0495 0496 0497 0498 0499 0500 0501 0502 0503 0504 0505 0506 0507 0508 0509 0510 0511 0512 0513 0514 0515 0516 0517 0518 0519 0520 0521 0522 0523 0524 0525 0526 0527 0528 0529 0530 0531 0532 0533 0534 0535 0536 0537 0538 0539 0540 0541 0542 0543 0544 0545 0546 0547 0548 0549 0550 0551 0552 0553 0554 0555 0556 0557 0558 0559 0560 0561 0562 0563 0564 0565 0566 0567 0568 0569 0570 0571 0572 0573 0574 0575 0576 0577 0578 0579 0580 0581 0582 0583 0584 0585 0586 0587 0588 0589 0590 0591 0592 0593 0594 0595 0596 0597 0598 0599 0600 0601 0602 0603 0604 0605 0606 0607 0608 0609 0610 0611 0612 0613 0614 0615 0616 0617 0618 0619 0620 0621 0622 0623 0624 0625 0626 0627 0628 0629 0630 0631 0632 0633 0634 0635 0636 0637 0638 0639 0640 0641 0642 0643 0644 0645 0646 0647 0648 0649 0650 0651 0652 0653 0654 0655 0656 0657 0658 0659 0660 0661 0662 0663 0664 0665 0666 0667 0668 0669 0670 0671 0672 0673 0674 0675 0676 0677 0678 0679 0680 0681 0682 0683 0684 0685 0686 0687 0688 0689 0690 0691 0692 0693 0694 0695 0696 0697 0698 0699 0700 0701 0702 0703 0704 0705 0706 0707 0708 0709 0710 0711 0712 0713 0714 0715 0716 0717 0718 0719 0720 0721 0722 0723 0724 0725 0726 0727 0728 0729 0730 0731 0732 0733 0734 0735 0736 0737 0738 0739 0740 0741 0742 0743 0744 0745 0746 0747 0748 0749 0750 0751 0752 0753 0754 0755 0756 0757 0758 0759 0760 0761 0762 0763 0764 0765 0766 0767 0768 0769 0770 0771 0772 0773 0774 0775 0776 0777 0778 0779 0780 0781 0782 0783 0784 0785 0786 0787 0788 0789 0790 0791 0792 0793 0794 0795 0796 0797 0798 0799 0800 0801 0802 0803 0804 0805 0806 0807 0808 0809 0810 0811 0812 0813 0814 0815 0816 0817 0818 0819 0820 0821 0822 0823 0824 0825 0826 0827 0828 0829 0830 0831 0832 0833 0834 0835 0836 0837 0838 0839 0840 0841 0842 0843 0844 0845 0846 0847 0848 0849 0850 0851 0852 0853 0854 0855 0856 0857 0858 0859 0860 0861 0862 0863 0864 0865 0866 0867 0868 0869 0870 0871 0872 0873 0874 0875 0876 0877 0878 0879 0880 0881 0882 0883 0884 0885 0886 0887 0888 0889 0890 0891 0892 0893 0894 0895 0896 0897 0898 0899 0900 0901 0902 0903 0904 0905 0906 0907 0908 0909 0910 0911 0912 0913 0914 0915 0916 0917 0918 0919 0920 0921 0922 0923 0924 0925 0926 0927 0928 0929 0930 0931 0932 0933 0934 0935 0936 0937 0938 0939 0940 0941 0942 0943 0944 0945 0946 0947 0948 0949 0950 0951 0952 0953 0954 0955 0956 0957 0958 0959 0960 0961 0962 0963 0964 0965 0966 0967 0968 0969 0970 0971 0972 0973 0974 0975 0976 0977 0978 0979 0980 0981 0982 0983 0984 0985 0986 0987 0988 0989 0990 0991 0992 0993 0994 0995 0996 0997 0998 0999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084

db 934 ---GCATCGTGGTGGAAACATGTCCTTCTGCTCTGCGAGTGGCCCTGGCAGGCCACTT 989

Oy	991	caagcgcgaagaacgctcaagctgtgaggagagcccaatacaacccccagtgatcgtgaca	990
Db	990	CAAGTCCAGGGCTACACCTGTGGGGGGCTGTCTACACAGCCCTGTGGATCATCACT	1049
Oy	991	gccgcacacgtcgctggaaaaaaccttaacaatccatgacattgaaagcattctgcggg	1050
Db	1050	GCTGCACACTGTGTTTAAGA--CTTGTACTCTCCCAAGTATGGA---CCATCCAGGTGG	1104
Oy	1051	attttgagacaactcttcaatgctctatgagcggatataccaagttagaaaaatgattct	1110
Db	1105	GTCAGTTCCTCCCTGTGGACAATCCA--GCCCAATCCCACTTGCGGAGAAAGATGTCTAC	1163
Oy	1111	catccaattatgactccaagaaccaagaacaatagaatctgctcgatgaagctgcagaag	1170
Db	1164	CACAGCAAGTCAAGCCMAAAGAGCTGGCCAAATACATTCGCCCTTATGAAGCTGGCCGG	1223
Oy	1171	cctctgaacttcaagagactagtgaatacaagtgtctctgcccacccagcaatgactg	1230
Db	1224	CCACTCACGTTCAATGAATATCCAGCCTTGTGCTTGCCCAACTTGAAGACAACTTC	1283
Oy	1231	cagccagaacacagctctgctgattcccggtgtygggggccaacgaga--gaaagggaa	1287
Db	1284	CCCGATGGAATAAGTCTGTGGACGTCAGAGATGGGGGCCACAGAGATGAGCAGGTGAC	1343
Oy	1288	accctcaagaatgctcgaaagctgcgaaggtgctctcatatgaacacagatgcacaagc	1347
Db	1344	GCTTCCTCCCTGTCTTGAAACACAGCGGCGCTCCCTTGTGATTTCCAAACAAGATCTGCACAC	1403
Oy	1348	agatagtctatgacaacctgaactgaacacagacaaatgaatctgtgcagctctctgcagg	1407
Db	1404	AGGACAGTGTACGGTGGGATCATCTCCCTTCATGCTCTGCGGGCTTACCTGACGGGT	1463
Oy	1408	aacgtcgattcttcccaaggtgacagatgaaaggccctctgtgtcaacttcgaagaacaatc	1467
Db	1464	GCGCGGACACTGCTGCAGAGGGGACAGCGGGGGCCCTCGGTGTCAAGAGAGAGGCTG	1523
Oy	1468	tgtgtggtcgataagggaatacaagaactgtgggtctctgctgtgcgaagcttcacagaac	1527
Db	1524	TGGAAATTAGTGGAGGACACACAGCTTTGGCATCGCGCCGAGAGGTGAACACCTGGGG	1583
Oy	1528	gtgtacgggaattgatatgtatctcaaggaactggaatttatcgacaatgaatgaaggcagc	1587
Db	1584	GTTGACACCGCTGTACCTCTTCTGTGACTGATCCACAGCAGATGAGAGAGACTTA	1643
Oy	1588	taatccacatg 1598	
Db	1644	AAAACCTCAG 1654	
RESULT	15		
LOCUS	AF201380	2416 bp mRNA	PRI 03-DEC-2000
DEFINITION	Homo sapiens serine protease TADG12 mRNA, complete cds.		
ACCESSION	AF201380		
VERSION	AF201380.1	GI:11527821	
KEYWORDS			
ORGANISM	human.		
REFERENCE			
AUTHORS	1 (bases 1 to 2416)		
TITLE	Underwood, L.J., Shigemasa, K., Tanimoto, H., Beard, J.B., Schneider, E.N., Wang, Y., Parmley, T.H. and O'Brien, T.J. Ovarian tumor cells express a novel multi-domain cell surface serine protease		
JOURNAL	Biochim. Biophys. Acta 1502 (3), 337-350 (2000)		
PUBMED	11068177		
REFERENCE	2 (bases 1 to 2416)		
AUTHORS	Underwood, L.J., Tanimoto, H., Schneider, E., Wang, Y., Parmley, T., Shigemasa, K. and O'Brien, T.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-NOV-1999) OBGYN, University of Arkansas for Medical Sciences, 4301 W. Markham St., Little Rock, AR 72205, USA		

FEATURES		Location/Qualifiers
Source	1..2416	/organism="Homo sapiens" /db_xref="taxon:9606" 144..1511 /codon_start=1 /product="serine protease TADG12" /protein_id="AAG31012.1" /db_xref="gi:11527822" /translation="MGENDPPAVPAEPFSRSLFGLDLKISPAVDADVAAQILSLPIPEEYRVSQSSLGIIALILALIGLIGIHDCGSKTKRNSFICILIRIDCDVSPCKDQEDKRCVRYGVGNOMAVLQVETPAISMKTGSDMKKRYANVACDLPFSYSNDLIRYSLEBOFRREPIYSIDHLLPDDKVPYLAHLHSYVIREGCASRHYVTLDOCTACGRKRSYIVGNNLSLSQMPQASLQFOGYHLGCGSVYLPMTITPAHCVOLYLPKSWTIDGVLVSLNDPAPSLVLEKIVYKSKPRKRLNDLMLMIAQPLFENMIQVCPNSENRPEKQKMTSGMGATEDGSDASPLVNLHAAVPLNNDLPOGVGRVHMLPLHALLRDPGGWRNMSQCGDSGGLVPCQERRLMKLVGATSPFICADVANKPGVYTRVYSLFDWIHEQNEKRLKPT"
CDS		
BASE COUNT	555 a 639 c 636 g 586 t	
ORIGIN		
Query Match	12.0%: Score 208.6; DB 89; Length 2416;	
Best Local Similarity	54.6%: Prod. No. 1.2e-44;	
Matches 628; Conservative	0; Mismatches 484; Indels 39; Gaps 9;	
469	lpgagaccctccagagagacccgcatcaaccctctaactggtgtagtgcgtgtcacactg	528
Db		
381	TCCTGCTCATCTTTAAGTGTATCGAGCGATACATCGATGTGACGAGATCGGATTGC	440
QY	cccgccggggagagacgagatcgtgtgtgttcgcctcttaagcaaaactcatccttaag	588
Db		
441	AAAGACGGGGAGAGACGAGATCCGCTGTGCCGGGTGGGTGCAGAAATGCCCTGCTCAG	500
QY	gtgtactcaatccagagaaagtccttcggcaaccctgtgtccaaagacagcttgaaagagac	648
Db		
501	GTGTTACACAGCTGCT-----TCGTGGAAGACCATGTGCTCCGATGACTGGAGAGTCA	554
QY	taagcgccggcgccctgcagagagacatggggctataaagaataatttactctlaagccaaaga	708
Db		
555	TACGCAAAATGTTGCTGTCGCCACACTGGGTTTCCAGAGCTATGTGATGATACAGATAACCTC	614
QY	atagtg-gatgaagcgatccacacagctttagaaactgaacacaaagtcgcggcaatgt	767
Db		
615	AGAGTGAGCTCCTCGAGGGGCGATTCGGGGAGGATTGTGTCATGATCACCCTTTCG	674
QY	gcatctcataaaaaactgttacac-----agtgatgcctgtctctca	810
Db		
675	CCAGATGACAAAGTACTGCATTACACCACTCAGTATATGTGAGGGAGAGATGTCTCTCT	734
QY	aaagatgagtgcttcttaacgctgataagcctgcggggtcaacttgaactcaagccggcaag	870
Db		
735	GGCCACAGGTTTACCTTGACAGTGCACAGCGTGTGTCATAGAAAGGGGCTACAGCTCAC--	792
QY	agcaagatgtgvgcgcgagagcgcgctcccggggacctggccctgcaggttgaagcctg	930
Db		
793	----GCATTCGTGGGTGGAACATGTCCTTGTGCTCTGCGAGTGGCCCTGCGACGCGCTT	848
QY	caagtcacgaagaagtcgaagtgctgacatcaatcaaaccccggaatgtagatgtagca	990
Db		
849	CAGTTCACAGGGGTACCACTGTGCGGGGGCTGTGTCATCAGCCCTCTGTGATCATCACT	908
QY	gcgcgcacactgcgtggaanaaacctcttaacaaatcaatgcatgtgaacgcatlttcggag	1050
Db		
909	GCTGACACTGTGTTATGA--CTTGTAACCTGCCAATGACATGGA---CCATCCAGATGG	963
QY	atttgagacaactcttcatgttctatgtaggcggatcaaccaagtagaanaagtgatctct	1110
Db		
964	GTCGTGTTCCCTGTGGACATCA--GCCCATCCCACTTGTTGGAAGAAGATTGTTAC	1022
QY	catccaatttagactccaagaccgaagaanaatgcatlttgctgtgataagctgcaagaag	1170
Db		
1033	CACACCAAGTACAAACCCAAAGAGGCTGGGCAATGACATGCGCCCTTATCAAGCTGGCCGG	1082

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 09:21:33 ; Search time 103.35 Seconds

10559.197 Million cell updates/sec

Title: US-09-615-285-1

Sequence: 1 ggcgagggcgagggcgag.....ctgcaaaaaaaaaa 1738

Scoring table: . IDENTITY_NUC

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters:	1460202
--	---------

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_0601:*

1:	/SIDS1/gcgdata/genseq/genseqn/NA1980.DAT *
2:	/SIDS1/gcgdata/genseq/genseqn/NA1981.DAT *
3:	/SIDS1/gcgdata/genseq/genseqn/NA1982.DAT *
4:	/SIDS1/gcgdata/genseq/genseqn/NA1983.DAT *
5:	/SIDS1/gcgdata/genseq/genseqn/NA1984.DAT *
6:	/SIDS1/gcgdata/genseq/genseqn/NA1985.DAT *
7:	/SIDS1/gcgdata/genseq/genseqn/NA1986.DAT *
8:	/SIDS1/gcgdata/genseq/genseqn/NA1987.DAT *
9:	/SIDS1/gcgdata/genseq/genseqn/NA1988.DAT *
10:	/SIDS1/gcgdata/genseq/genseqn/NA1989.DAT *
11:	/SIDS1/gcgdata/genseq/genseqn/NA1990.DAT *
12:	/SIDS1/gcgdata/genseq/genseqn/NA1991.DAT *
13:	/SIDS1/gcgdata/genseq/genseqn/NA1992.DAT *
14:	/SIDS1/gcgdata/genseq/genseqn/NA1993.DAT *
15:	/SIDS1/gcgdata/genseq/genseqn/NA1994.DAT *
16:	/SIDS1/gcgdata/genseq/genseqn/NA1995.DAT *
17:	/SIDS1/gcgdata/genseq/genseqn/NA1996.DAT *
18:	/SIDS1/gcgdata/genseq/genseqn/NA1997.DAT *
19:	/SIDS1/gcgdata/genseq/genseqn/NA1998.DAT *
20:	/SIDS1/gcgdata/genseq/genseqn/NA1999.DAT *
21:	/SIDS1/gcgdata/genseq/genseqn/NA2000.DAT *
22:	/SIDS1/gcgdata/genseq/genseqn/NA2001.DAT *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1738	100.0	1738	21	AAZ29368	Human 201P12-GTC
2	1721.4	99.0	3245	21	AAZ08803	Androgen-inducible
3	1710.4	98.4	3443	21	AAZ995005	Cancer specific ge
4	1655.2	95.2	2479	21	AAZ90478	Ovrl15 homolog pro
5	1655.2	95.2	2479	21	AAZ87813	Human tumour supp
6	1655.2	95.2	3966	21	AAZ83325	Human TMRPSS2 DNA
7	1472.4	84.7	1479	21	AAZ87785	Human tumour supp
8	992.4	57.1	1076	21	AAAI2975	cdNA encoding huma
9	267	15.4	267	21	AAZ95504	CSG Prol15 EST use
10	249	14.3	272	21	AAAC06945	Human secreted pro
11	230.4	13.3	318	21	AAAC98460	Human colon cancer

[illegible]

ALIGNMENTS

RESULT	1
AAZ29636	
ID	AAZ29636 standard; cDNA, 1738 BP.
XX	
AC	AAZ29636;
XX	
DT	22-MAR-2000 (first entry)
XX	
DE	Human 20P1F12-GTC2 encoding cDNA.
XX	
KW	20P1F12; TMRSS2; androgen; serine protease; 20P1F12-GTC1; cancer
KV	transmembrane protein; colon; prostate; prostate tumour; ss.
XX	
OS	Homo sapiens.
XX	
EH	Location/Qualifiers
FT	112..1590
CDS	
FT	/*tag= a
FT	/product= "20P1F12-GTC2"
FT	/note= "Prostate-specific serine protease"
FN	W09962942-A2.
XX	
PD	09-DEC-1999.
XX	
PF	01-JUN-1999; 99WO-US12253.
XX	
PR	01-JUN-1998; 98US-0087598.
PR	29-JUN-1998; 98US-0091474.
PR	14-APR-1999; 99US-0129521.
XX	
PA	(UROS-) UROGENESYS INC.
PA	(AFAR-) AFAR D E.
DA	(HUBE/) HUBERT R S.


```

|||||
DB 758 agcgatccaccagcttatgaactgaacacaaagtcgagcaatgctgatatcataaa 817
OY 781 aaactgtaccacagtgatgctgttcttcaaaagcagtggttcttcaagctatagcc 840
DB 818 aaactgtaccacagtgatgctgttcttcaaaagcagtggttcttcaagctatagcc 877
OY 841 tggcggtgtaactgtgaactcaagccgacagagagatgttgcgcgagcgagcgctc 900
DB 878 tggcggtgtaactgtgaactcaagccgacagagagatgttgcgcgagcgagcgctc 937
OY 901 ccggtgggcttgccctt-ggcaaggtcagcctgcagctccacagaacgtccacgtgtcgagag 959
DB 938 ccggtgggcttgcccttgagaggttcagccgcagctccacagaacgtccacgtgtcgagag 997
OY 960 ctccatcatcaccctccagatgagatcgttgacagccgcccaactcgtgtgaaaaacctttaa 1019
DB 998 ctccatcatcaccctccagatgagatcgttgacagccgcccaactcgtgtgaaaaacctttaa 1057
OY 1020 caatcatcagcatctgagcagcagcttcgagagatttgagacaactcttctatgtctatg 1079
DB 1058 caatcatcagcatctgagcagcagcttcgagagatttgagacaactcttctatgtctatg 1117
OY 1080 agccggaataccaaagtgaanaaagtgtatctcatalccaattatgactccaagaaccaagaa 1139
DB 1118 agccggaataccaaagtgaanaaagtgtatctcatalccaattatgactccaagaaccaagaa 1177
OY 1140 caatgacatctgagcagtgatgaagctgcagaagcctctgacttcaagaactatgtgaacc 1199
DB 1178 caatgacatctgagcagtgatgaagctgcagaagcctctgacttcaagaactatgtgaacc 1237
OY 1200 aggtgtctctcccaaccagcagtgatgtctgcagccagaacagctcgtcggtttccg 1259
DB 1238 aggtgtctctcccaaccagcagtgatgtctgcagccagaacagctcgtcggtttccg 1297
OY 1260 gtggggggccacccagagagaaaggaagacactcaagaagtgtcgaagcgtgccaaagtgtct 1319
DB 1298 gtggggggccacccagagagaaaggaagacactcaagaagtgtcgaagcgtgccaaagtgtct 1357
OY 1320 tctcatttgagacacagatagacagcagcagatgtctatagaacactgtctacacagc 1379
DB 1358 tctcatttgagacacagatagacagcagcagatgtctatagaacactgtctacacagc 1417
OY 1380 catgactgtgcccgtctcctcaggggaagcgtcgaattctgcaggggtgacagtgagag 1439
DB 1418 catgactgtgcccgtctcctcaggggaagcgtcgaattctgcaggggtgacagtgagag 1477
OY 1440 gctctggtcactctgaaagacaatatctgtgtgctgatataggggatacaagctggggttc 1499
DB 1478 gctctggtcactctgaaagacaatatctgtgtgctgatataggggatacaagctggggttc 1537
OY 1500 tggctgtgcacaaagcttaccagacagagatgtacaggaagtgtatgttatccagcagctg 1559
DB 1538 tggctgtgcacaaagcttaccagacagagatgtacaggaagtgtatgttatccagcagctg 1597
OY 1560 gatttctgcacaaatagggcagacagcgtacatccaaatggtcttccttgcagcgttc 1619
DB 1598 gatttctgcacaaatagggcagacagcgtacatccaaatggtcttccttgcagcgttc 1657
OY 1620 ttacaagaanaaaataggggagctgttctgtctcccgtagatgatttaactttagaagatga 1679
DB 1658 ttacaagaanaaaataggggagctgttctgtctcccgtagatgatttaactttagaagatga 1717
OY 1680 ttcaagaggtcactcttatttataaagctgtaactgtgtctg 1723
DB 1718 ttcaagaggtcactcttatttataaagctgtaactgtgtctg 1761

```

RESULT 4
 AA290478
 ID AA290478 standard; cDNA; 2479 BP.
 XX
 AC AA290478;

```

XX 06-JUN-2000 (first entry)
DT
XX
DE Ovr115 homolog protein encoding cDNA.
XX
KW CSG; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;
XX endometrial; uterine; lung; cytotoxic; ss.
XX
OS Homo sapiens.
XX
PN WO200012758-A1.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US19655.
XX
PR 02-SEP-1998; 98US-0098880.
XX
PA (DIAD-) DIADEXUS LLC.
XX
PI Salceda S, Sun Y, Recipon H, Caferkey R;
XX
DR WPI; 2000-256657/22.
DR
DR P-PSDB; AAY57280.
XX
PT Diagnosing, staging, monitoring, imaging and treating cancer especially
PT gynecological cancers e.g. breast, ovarian cancer and lung cancer,
PT involves measuring cancer specific gene levels in cells and body fluids
XX
PS
XX
XX Claim 9; Page 49-50; 58pp; English.
XX
CC The invention relates to detecting, diagnosing metastasis and staging
CC cancer by measuring levels of cancer specific genes (CSG) in cells,
CC tissues or body fluids. Their remission and progression, decreases and
CC increases in CSG levels, is also monitored, by periodic sample analysis.
CC The methods are useful for detecting cancers, especially gynecologic
CC cancers which include ovarian, breast, endometrial and uterine cancer
CC and lung cancer. Antibodies against the CSGs labeled with paramagnetic
CC ions or a radioisotope is useful for imaging cancer and when conjugated
CC with a cytotoxic agent are useful for treating cancer. The present
CC sequence represents a Ovr115 homolog protein encoding cDNA, that
CC can be used for the detection of the various cancers.
XX
XX
SQ Sequence 2479 BP; 578 A; 650 C; 677 G; 574 T; 0 other;

```

Query Match 95.2%; Score 1655.2; DB 21; Length 2479;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

OY 56 gtacatatgaacattccagatacctacatcattactcagctgtgtgataacagcaagatg 115
DB 1 gtcattatgacattccagatacctacatcattactcagctgtgtgataacagcaagatg 60
OY 116 cttagaactaaggttaacacacagctatctgacctactatgaaacacatgatatcaaac 175
DB 61 cttagaactaaggttaacacacagctatctgacctactatgaaacacatgatatcaaac 120
OY 176 cggaaacccctatccgcacagccacactgtgtgtcccaactgtctacagaggtgcatccg 235
DB 121 cggaaacccctatccgcacagccacactgtgtgtcccaactgtctacagaggtgcatccg 180
OY 236 ctgaagtactaccgttcccggtgcccagatgaagcccgaggtctcgaagcaggtcttca 295
DB 181 ctgaagtactaccgttcccggtgcccagatgaagcccgaggtctcgaagcaggtcttca 240
OY 296 accccgtgctgcagcagcccaaatcccatcccgagagctgtagcacttaaaagacta 355
DB 241 accccgtgctgcagcagcccaaatcccatcccgagagctgtagcacttaaaagacta 300
OY 356 agaagacgtgtgcatcacttgaccctggagaccttccctctgtgagagctgctgtccg 415

```

Dd 301 agaaagcactgtgcatcacccttgagcccttgaggacccttcctcgttgaggactggtccg 360
Oy 416 cggccactctcggaaagtcatcaggcagaagtgctccaacctgggacagatgagct 475
Dd 361 cgggctactctcggaaagtcatcaggcagaagtgctccaacctgggacagatgagct 420
Oy 476 cctcaggtacatcgtcaaacccctcctaactgtgtatgtcggtgtcaacatgccccg 535
Dd 421 cctcaggtacatcgtcaaacccctcctaactgtgtatgtcggtgtcaacatgccccg 480
Oy 536 gggaggaagagaatcgt 595
Dd 481 gggaggaagagaatcgt 540
Oy 596 cctccagaagagaatcgt 655
Dd 541 cctccagaagagaatcgt 600
Oy 656 gggcggcctcgaaggagcatgggtctatagaataatttctaactctagccaaggaatgtg 715
Dd 601 gggcggcctcgaaggagcatgggtctatagaataatttctaactctagccaaggaatgtg 660
Oy 716 atgacagcggatccaccagcttatagaatactgaacacaaagtgcggcaatgtcgaatct 775
Dd 661 atgacagcggatccaccagcttatagaatactgaacacaaagtgcggcaatgtcgaatct 720
Oy 776 ataaataaacctgtacacaaagtgaatgacctgtcttcaaaaagcagtggtcttcaagctga 835
Dd 721 ataaataaacctgtacacaaagtgaatgacctgtcttcaaaaagcagtggtcttcaagctga 780
Oy 836 tagcctcgggggtcacaacttgaaactcaagcgcgcagagcaggaatgtgtggcggcagagc 895
Dd 781 tagcctcgggggtcacaacttgaaactcaagcgcgcagagcaggaatgtgtggcggcagagc 840
Oy 896 cgtcccgaggggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcgg 955
Dd 841 cgtcccgaggggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcgg 900
Oy 956 gagggtccatcacaaccccgagtgatcgtgacagcgcgcacatgctgctgagaaaaaccc 1015
Dd 901 gagggtccatcacaaccccgagtgatcgtgacagcgcgcacatgctgctgagaaaaaccc 960
Oy 1016 ttaacaaatccatggaatcgtgacagcgtctgaggagattttgagaaatcttcaatgtct 1075
Dd 961 ttaacaaatccatggaatcgtgacagcgtctgaggagattttgagaaatcttcaatgtct 1020
Oy 1076 atgggagccggatccaaagtagaaaaagtgattctcattccaaatcttgactccaagaca 1135
Dd 1021 atgggagccggatccaaagtagaaaaagtgattctcattccaaatcttgactccaagaca 1080
Oy 1136 agaaacaatgacaatgctgactgataagctgacagaagcctctgacttcaacgaactatgta 1195
Dd 1081 agaaacaatgacaatgctgactgataagctgacagaagcctctgacttcaacgaactatgta 1140
Oy 1196 aaccagctgtgtctgcccacacacagagcatgactgacgcacaaacagctctgtgattt 1255
Dd 1141 aaccagctgtgtctgcccacacacagagcatgactgacgcacaaacagctctgtgattt 1200
Oy 1256 cggggtggggggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcgg 1315
Dd 1201 cggggtggggggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcgg 1260
Oy 1316 tgcctctcatctgagacaagagatgacaacagagatgatactcaaacacgcatcacac 1375
Dd 1261 tgcctctcatctgagacaagagatgacaacagagatgatactcaaacacgcatcacac 1320
Oy 1376 cagccatgatactgtgcccgtctcctgcagggagacgtgattctctgcagggtgacagtg 1435
Dd 1321 cagccatgatactgtgcccgtctcctgcagggagacgtgattctctgcagggtgacagtg 1380
Oy 1436 gagggcctctgtgatactcctcgaagaaacaatactgtgtgtgtgtgtgtgtgtgtgtgtgt 1495
Dd 1381 gagggcctctgtgatactcctcgaagaaacaatactgtgtgtgtgtgtgtgtgtgtgtgtgt 1440

Oy 1496 gtctcgtcgtgtccaaagcttaacagcagaggtgtacgggaatgtatgtatcacgg 1555
Dd 1441 gtctcgtcgtgtccaaagcttaacagcagaggtgtacgggaatgtatgtatcacgg 1500
Oy 1556 actggaattatcgaacaatgagggcagacggtcaatccacatgctctcgtcttgagct 1615
Dd 1501 actggaattatcgaacaatgagggcagacggtcaatccacatgctctcgtcttgagct 1560
Oy 1616 cgtttacaagaagaacaatgagggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1675
Dd 1561 cgtttacaagaagaacaatgagggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1620
Oy 1676 atgatacgaaggctcacttcatcttatttaaaagtgaaactgtctggc 1723
Dd 1621 atgatacgaaggctcacttcatcttatttaaaagtgaaactgtctggc 1668

RESULT 5
AAZ87813
ID AAZ87813 standard; DNA; 2479 BP.
XX
AC AAZ87813;
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMRSS2 gene (GenBank Accn No: U75329).
XX
KW Tumour suppressor gene; TMRSS2; cancer; human; drug design;
KW gene therapy; protein therapy; ds.
XX
OS Homo sapiens.
XX
PN WO20000605-A1.
XX
PD 06-JAN-2000.
XX
PF 29-JUN-1999; 99WO-US14622.
XX
PR 29-JUN-1998; 98US-0091044.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Wong AKC, Tavligian SV, Teng DHF;
XX
DR WPI: 2000-170914/15.
XX
PT Novel tumor suppressor TMRSS2 used for the diagnosis and prognosis of
PI human cancer -
XX
PS Example 1; Page 84-85; 89pp; English.
XX
CC The invention provides a new tumour suppressor gene, designated TMRSS2.
CC The TMRSS2 polynucleotides and polypeptides can be used in methods for
CC diagnosing and prognosing predisposition to cancer in humans. The
CC polypeptides may also be used in assays to screen for compounds with
CC anti-cancer or therapeutic properties. The polypeptides are also useful
CC for rational drug design. The TMRSS2 polynucleotides and polypeptides
CC may be used for gene therapy and protein therapy. The present sequence
CC represents the TMRSS2 gene (Genbank Accn No: U75329).
XX
SQ Sequence 2479 BP; 578 A; 650 C; 677 G; 574 T; 0 other;

Query Match 95.2%; Score 1655.2; DB 21; Length 2479;
Best Local Similarity 99.5%; Pred. No. 0; Mismatches 8; Indels 0; Gaps 0;
Matches 1660; Conservative 0;

Oy 56 gtcatatgaaatccagatacctatcatctcgtatgtctgtataacagcaagatg 115
Dd 1 gtcatatgaaatccagatacctatcatctcgtatgtctgtataacagcaagatg 60
Oy 116 ctltgaaactcagggtccaccacagctatgtgaccttactactgataaacaatgataccaac 175

Db 61 cttgagactcagggtcaccacacagctatgtgacctactactatgaataacatgatalcaaac 120
QY 176 cggaaaaaccctatcccgacagccacatgtgtgtcccccactgtctcagagtgatcccg 235
Db 121 cggaaaaaccctatcccgacagccacatgtgtgtcccccactgtctcagagtgatcccg 180
QY 236 ctgagactacacccgtcccccgtgtcccccagtagcccccaggggtccctgagccagagcttcca 255
Db 181 ctgagactacacccgtcccccgtgtcccccagtagcccccaggggtccctgagccagagcttcca 240
QY 296 accccgtctgtcagacgagccccaatcccccacccgagagctgtcagctcacaagacta 355
Db 241 accccgtctgtcagacgagccccaatcccccacccgagagctgtcagctcacaagacta 300
QY 356 agaaagcaactgtgcatcaacttgaacctgtggagaccttcctgtggagctgtgcgtgcg 415
Db 301 agaaagcaactgtgcatcaacttgaacctgtggagaccttcctgtggagctgtgcgtgcg 360
QY 416 ctgagctactctggaagtctcagagcaagtgtctccaaactctggagatagtgagct 475
Db 361 ctgagctactctggaagtctcagagcaagtgtctccaaactctggagatagtgagct 420
QY 476 cctcaggtactcgtcatcaacccctctaaactgtgtgtatgtggtgtcacaactgtcccgcg 535
Db 421 cctcaggtactcgtcatcaacccctctaaactgtgtgtatgtggtgtcacaactgtcccgcg 480
QY 536 gggaggaagcagaatcgt 555
Db 481 gggaggaagcagaatcgt 540
QY 596 catctcagaggaagtctgt 655
Db 541 catctcagaggaagtctgt 600
QY 656 gggcggtcgtcagagcagatgt 715
Db 601 gggcggtcgtcagagcagatgt 660
QY 716 atgacagcggatccacagcttataatgaacacaaagtgtccggcaatgtctgatatct 775
Db 661 atgacagcggatccacagcttataatgaacacaaagtgtccggcaatgtctgatatct 720
QY 776 ataaaaactgtacacacagctgt 835
Db 721 ataaaaactgtacacacagctgt 780
QY 836 taagctcgtcgt 895
Db 781 taagctcgtcgt 840
QY 896 cgtctcccggt 955
Db 841 cgtctcccggt 900
QY 956 gaggtcctcatcaccaccccgagtgatgtgacagccgcccactcgtgtgtgtgtgtgtgtgtgt 1015
Db 901 gaggtcctcatcaccaccccgagtgatgtgacagccgcccactcgtgtgtgtgtgtgtgtgtgt 960
QY 1016 ttaacaatcatgtgcatcttgagcggcatcttgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1075
Db 961 ttaacaatcatgtgcatcttgagcggcatcttgcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
QY 1076 atggggtcggatataccaaagtagaagaaagtattctcatccaattatgtatccaagacca 1135
Db 1021 atggggtcggatataccaaagtagaagaaagtattctcatccaattatgtatccaagacca 1080
QY 1136 agaaacatgacattgtcgt 1195
Db 1081 agaaacatgacattgtcgt 1140
QY 1196 aaccagtgctgt 1255

Db 1141 aaccagtgctgt 1200
QY 1256 cgggt 1315
Db 1201 cgggt 1260
QY 1316 tgccttcatcttgagaaacagagatgtcacagcagcagatgtgtctatgaacactgtatcacac 1375
Db 1261 tgccttcatcttgagaaacagagatgtcacagcagcagatgtgtctatgaacactgtatcacac 1320
QY 1376 cagccatgacatctgtccgggtcttcctcagaggaagcgtcagatctcttcgacaggtgtgacagtg 1435
Db 1321 cagccatgacatctgtccgggtcttcctcagaggaagcgtcagatctcttcgacaggtgtgacagtg 1380
QY 1436 gagggcctcgtgtcactctgaaagacaatatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1495
Db 1381 gagggcctcgtgtcactctgaaagacaatatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
QY 1496 gtctgt 1555
Db 1441 gtctgt 1500
QY 1556 actggaattatcgaacaaatgaaggcagacgctaatacagatgtctcgtctcgttgaact 1615
Db 1501 actggaattatcgaacaaatgaaggcagacgctaatacagatgtctcgtctcgttgaact 1560
QY 1616 cgttttcaagaagaacaatgt 1675
Db 1561 cgttttcaagaagaacaatgt 1620
QY 1676 atgattcagaggtcacttcttatttataaagatgaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1723
Db 1621 atgattcagaggtcacttcttatttataaagatgaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1668

RESULT 6
AAC83325
ID AAC83325 standard; DNM; 3966 BP.
XX
AC AAC83325;
XX
DT 26-FEB-2001 (first entry)
XX
DE Human TMPRSS2 DNA.
XX
KW Prostate specific androgen regulated protein; ARSDRL; TMPRSS2;
KW PART-1; neoplastic; ds.
OS Homo sapiens.
PN MO200065067-A2.
XX
PD 02-NOV-2000.
XX
PF 21-APR-2000; 2000MO-US10920.
XX
PR 23-APR-1999; 99US-0130778.
PR 30-AUG-1999; 99US-0151585.
PR 30-DEC-1999; 99US-0174003.
PR 24-JAN-2000; 2000US-0177751.
XX
XX
XX (UNIM) UNIV WASHINGTON.
XX
XX
XX PI Nelson PS, Hood L, Lin B;
XX
XX DR MPI: 2000-679676/66.
XX DR P-PSDB; AAB36901.
XX
XX PT Polynucleotide encoding prostate specific androgen regulated
XX PT polypeptides and inhibitor of the peptides useful for treating or
XX PT reducing the progression of prostate neoplastic condition in an
XX PT individual -

PS Claim 1; Page 81-86; 121pp; English.
XX The present invention relates to prostate specific androgen regulated
CC proteins. The invention may be used to determine an expression level
CC of the prostate-specific proteins ARSDR1, TMPRSS2, or PAPF-1 in a
CC fluid sample or prostate cell sample from an individual. It may also
CC be used for diagnosing and predicting the susceptibility of a
CC prostate neoplastic condition in an individual. Inhibitors of the
CC proteins are useful for treating or preventing the progression of a
CC prostate neoplastic condition.
XX
SQ Sequence 3966 BP; 992 A; 998 C; 988 G; 984 T; 4 other:

Query Match 95.2%; Score 1655.2; DB 21; Length 3966;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 56 gtcataatgaaatccagatccatataatcattacatgagctgataaagcaagatg 115
DB 1 gtcataatgaaatccagatccatataatcattacatgagctgataaagcaagatg 60
QY 116 ctttgaacctgaggtcaccacagcagctatgagccttactatgaataaaccatgataccaac 175
DB 61 ctttgaacctgaggtcaccacagcagctatgagccttactatgaataaaccatgataccaac 120
QY 176 cggaaaaccccatcccgacacagccacactgtgtcccaactgtctacagaggtgacatccg 235
DB 121 cggaaaaccccatcccgacacagccacactgtgtgtcccaactgtctacagaggtgacatccg 180
QY 236 ctcaatgataccgctcccgctgcccagtaacgcccagaggttccttgacgacaggttcca 295
DB 181 ctcaatgataccgctcccgctgcccagtaacgcccagaggttccttgacgacaggttcca 240
QY 296 accccgctgctgcagcagcagcccaaatcccaatcccgagagctgagacatcaagaagacta 355
DB 241 accccgctgctgcagcagcagcccaaatcccaatcccgagagctgagacatcaagaagacta 300
QY 356 agaaagcaactgtgcatcaaccttgagaccttgagaccttcctctgaggagctgagctgagc 415
DB 301 agaaagcaactgtgcatcaaccttgagaccttgagaccttcctctgaggagctgagctgagc 360
QY 416 ctgagcctactctggaagtctcatgtggcagcaagtgctccaacctgtggataaggttcgact 475
DB 361 ctgagcctactctggaagtctcatgtggcagcaagtgctccaacctgtggataaggttcgact 420
QY 476 cctcaggtacactgcatcaacctctaaactgtgtgtagtgctgtcacaacctgcccggcg 535
DB 421 cctcaggtacactgcatcaacctctaaactgtgtgtagtgctgtcacaacctgcccggcg 480
QY 536 gggagagcaagaaatcagtgctgctcctctacagcaaaactctccttcagagtgact 595
DB 481 gggagagcaagaaatcagtgctgctcctctacagcaaaactctccttcagagtgact 540
QY 596 catctcaagaggaagtcctcgtgcaacctgtgtgccaagaagactctgaaacgaatacagagc 655
DB 541 catctcaagaggaagtcctcgtgcaacctgtgtgccaagaagactctgaaacgaatacagagc 600
QY 656 gggcgagcctgcaagagcaatgaggtctaaagaataattttactctagccaagaaatagatg 715
DB 601 gggcgagcctgcaagagcaatgaggtctaaagaataattttactctagccaagaaatagatg 660
QY 716 atgacagcagatccacacagcttatagaatctgaacaagtgccgcaatgtcgaatct 775
DB 661 atgacagcagatccacacagcttatagaatctgaacaagtgccgcaatgtcgaatct 720
QY 776 ataaaaaactgtacacacagtgatgtccttctcaaaagcagtggtcttcttaagctgtta 835
DB 721 ataaaaaactgtacacacagtgatgtccttctcaaaagcagtggtcttcttaagctgtt 780
QY 836 tagcctgaggggtcaacttgaaactcaagccgacagagaggttctggcgcgacagagcg 895
DB 781 tagcctgaggggtcaacttgaaactcaagccgacagagaggttctggcggtgagagcg 840

QY 896 cgtcccgaggggcctgagcccttgacaggttcagcttcacagatcccaagatccacgtgtg 955
DB 841 cgtcccgaggggcctgagcccttgacaggttcagcttcacagatcccaagatccacgtgtg 900
QY 956 gaggctcatcatcaacccccagagtgatctgctgacagccgcccactcgtgtgaaaaactc 1015
DB 901 gaggctcatcatcaacccccagagtgatctgctgacagccgcccactcgtgtgaaaaactc 960
QY 1016 ttaacaatccatgagatctgagcagcatttcgagagattttgagacaatcttcatgtct 1075
DB 961 ttaacaatccatgagatctgagcagcatttcgagagattttgagacaatcttcatgtct 1020
QY 1076 atgagccgagataccacaaatgaaatgatttctcatccaaatataatgactccaagacca 1135
DB 1021 atgagccgagataccacaaatgaaatgatttctcatccaaatataatgactccaagacca 1080
QY 1136 agaaacaatgacatctgcgtgtgaaagctgacagaagcctctgacttcaacgaactagta 1195
DB 1081 agaaacaatgacatctgcgtgtgaaagctgacagaagcctctgacttcaacgaactagta 1140
QY 1196 aaccagtgctgtgcccacacccagcatgattgtcagcccaagacgctctgactt 1255
DB 1141 aaccagtgctgtgcccacacccagcatgattgtcagcccaagacgctctgactt 1200
QY 1256 ccgggttgaggggccacccagagagaaaggaaacactcagaagtgtctgaacgtctgcaagg 1315
DB 1201 ccgggttgaggggccacccagagagaaaggaaacactcagaagtgtctgaacgtctgcaagg 1260
QY 1316 tgccttctcatatgagacacaaatgacaaagatagctacagatagctatgatacaaacctgatacac 1375
DB 1261 tgccttctcatatgagacacaaatgacaaagatagctacagatagctatgatacaaacctgatacac 1320
QY 1376 cagccatgactgtgctgcggtctccctgcagggagaaagctcagatctctgcagaggtgacaatg 1435
DB 1321 cagccatgactgtgctgcggtctccctgcagggagaaagctcagatctctgcagaggtgacaatg 1380
QY 1436 gaggagcctctgtgtcaacttcgaagaacaatatctgtgtgctgataaggtatataaagctgg 1495
DB 1381 gaggagcctctgtgtcaacttcgaagaacaatatctgtgtgctgataaggtatataaagctgg 1440
QY 1496 gttctgagctgtgcccagaagcttacaagcagaggtgataagagatgtatgataatcagag 1555
DB 1441 gttctgagctgtgcccagaagcttacaagcagaggtgataagagatgtatgataatcagag 1500
QY 1556 actgagttatcagacaatatgaaagcagagcgtataatccacatgagcttcgtctcagact 1615
DB 1501 actgagttatcagacaatatgaaagcagagcgtataatccacatgagcttcgtctcagact 1560
QY 1616 cgtttacaagaagaacaatgaggtgtgttctgcttcccgatgcatgattactcttagag 1675
DB 1561 cgtttacaagaagaacaatgaggtgtgttctgcttcccgatgcatgattactcttagag 1620
QY 1676 atgattcagaggttcaattctatttataaacaagtgaaactgtctggc 1723
DB 1621 atgattcagaggttcaattctatttataaacaagtgaaactgtctggc 1668

RESULT 7
AA287786
ID AA287786 standard; DNA; 1479 BP.
XX
AC AA287786;
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMPRSS2 encoding DNA.
XX
KW Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
KW gene therapy; protein therapy; ds.
OS Homo sapiens.
XX

FH Key Location/Qualifiers
FT CDS 1..1479
FT /*tag= a
FT /product= "TMPRSS2"
FT allele 478
FT /*tag= b
FT /note= "this base can be G or A with G being the more common allele. The codon will change from Val to Met"
FT allele 625
FT /*tag= c
FT /note= "this base can be T or A with T being the more common allele. The codon will change from Phe to Ile"
FT conflict 724
FT /*tag= d
FT /note= "listed as T in GenBank Accn No: U75329"
FT allele 768
FT /*tag= e
FT /note= "this base can be C or T with C being the more common allele. This is a silent polymorphism"
FT allele 777
FT /*tag= f
FT /note= "this base can be C or T with C being the more common allele. The codon is unaffected with both alleles encoding Gly"
FT allele 834
FT /*tag= g
FT /note= "this base can be C or T with C being the more common allele. This is a silent polymorphism"
FT conflict 985
FT /*tag= h
FT /note= "listed as C in GenBank Accn No: U75329"
FT conflict 1347
FT /*tag= i
FT /note= "listed as C in GenBank Accn No: U75329"
FT conflict 1466
FT /*tag= j
FT /note= "listed as A in GenBank Accn No: U75329"
FT conflict 1471
FT /*tag= k
FT /note= "listed as A in GenBank Accn No: U75329"
PN MO20000605-A1.
XX 06-JAN-2000.
XX PD
XX 29-JUN-1999; 99WO-US14622.
XX PF
XX 29-JUN-1998; 98US-0091044.
XX PR
XX (MYRI-) MYRIAD GENETICS INC.
XX PA
XX Wong AKC, Tavtigian SV, Teng DHF;
XX PI
XX WPI: 2000-170914/15.
XX DR P-PSDB; AAV77726.
XX DR
XX
PT Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of human cancer -
XX human cancer -
XX
PS Claim 22; Page 77-79; 89pp; English.
XX
CC The invention provides a new tumour suppressor gene, designated TMPRSS2.
CC The TMPRSS2 polynucleotides and polypeptides can be used in methods for
CC diagnosing and prognosing predisposition to cancer in humans. The
CC polypeptides may also be used in assays to screen for compounds with
CC anti-cancer or therapeutic properties. The polypeptides are also useful
CC for rational drug design. The TMPRSS2 polynucleotides and polypeptides
CC may be used for gene therapy and protein therapy. The present sequence
CC represents the DNA encoding the TMPRSS2 polypeptide.

XX
SQ Sequence 1479 BP; 360 A; 406 C; 402 G; 306 T; 5 other:

Query Match 84.7%; Score 1472.4; DB 21; Length 1479;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 112 atgctttgaactcaaggtgtacccacagctatttgacacctactatataaacaatgatatac 171
DB 1 atgctttgaactcaaggtgtacccacagctatttgacacctactatataaacaatgatatac 60
QY 172 caaccgaaaccccatcccgacagccacatggtgtccacacttcaagagtgcat 221
DB 61 caaccgaaaccccatcccgacagccacatggtgtccacacttcaagagtgcat 120
QY 232 ccgctcagctactaccgcttcccgctgcccagtaacgcccggaggttcctgaagcagct 291
DB 121 ccgctcagctactaccgcttcccgctgcccagtaacgcccggaggttcctgaagcagct 180
QY 292 tccaacccgctgtctgcacgacgacccaaatcccatccggagcagtggtacactaaag 351
DB 181 tccaacccgctgtctgcacgacgacccaaatcccatccggagcagtggtacactaaag 240
QY 352 actaagaacacactgtgcatcaccttgacaccttgaggacctctctgtggaagctgcgtg 411
DB 241 mmmmgaaacactgtgcatcaccttgacaccttgaggacctctctgtggaagctgcgtg 300
QY 412 gccgctgacctcttgcgaagtcatgtggcgaagaatgtgctccaactctggatagaagtg 471
DB 301 gccgctgacctcttgcgaagtcatgtggcgaagaatgtgctccaactctggatagaagtg 360
QY 472 gactcctcaagtgactctgcacatcaacccctcaactgtgtgatatggtgtcacaactccc 531
DB 361 gactcctcaagtgactctgcacatcaacccctcaactgtgtgatatggtgtcacaactccc 420
QY 532 ggcggggaagcagagatcgtgtgttcgctcttaaggacaaactcaactccttcagtg 591
DB 421 ggcggggaagcagagatcgtgtgttcgctcttaaggacaaactcaactccttcagtg 480
QY 592 tactcatctcaagaggaagtgcttgcacccctgtgtgcgaagcagctggaagaaactac 651
DB 481 tactcatctcaagaggaagtgcttgcacccctgtgtgcgaagcagctggaagaaactac 540
QY 652 ggcggggaagcagcctgcagaggaatggcgtatagaataatttttaactctcgccaaggaata 711
DB 541 ggcggggaagcagcctgcagaggaatggcgtatagaataatttttaactctcgccaaggaata 600
QY 712 gtggatgacagcgagatccacacagcttataagaactgaacaagaatgcccgaatgtcgat 771
DB 601 gtggatgacagcgagatccacacagcttataagaactgaacaagaatgcccgaatgtcgat 660
QY 772 atctataaaaactgttaccacagtgatgctgttcttcaaaaagcagtggtttcttaagc 831
DB 661 atctataaaaactgttaccacagtgatgctgttcttcaaaaagcagtggtttcttaagc 720
QY 832 tgtatagctctgggggttcaacttgaactcaagcggccaagcaggtttgtggcggcgag 891
DB 721 tgtatagctctgggggttcaacttgaactcaagcggccaagcaggtttgtggcggcgag 780
QY 892 agcgcgctccggggggcctcgccctgcaggtcagctcagctcagctcagaagcgtccagctg 951
DB 781 agcgcgctccggggggcctcgccctgcaggtcagctcagctcagctcagaagcgtccagctg 840
QY 952 tgcggagctccatcatcatcccccagtgatcgtgacagcgccacactcgtgtgaaaaa 1011
DB 841 tgcggagctccatcatcatcccccagtgatcgtgacagcgccacactcgtgtgaaaaa 900
QY 1012 cctcttaaatcatcagatctggaagcagcatctggcgggattttggaaacttttcag 1071
DB 901 cctcttaaatcatcagatctggaagcagcatctggcgggattttggaaacttttcag 960
QY 1072 ttctatggagccgatacatcaagaatagaatgtatttccatccaatattgactccaag 1131


```
Db 682 ctaattgagacacagatgtaacacagatalgtctatgacaacctgatacacacagcc 741
QY 1381 atgattcgttcgggtctccgacggggaacgttgatctctgccaagggtgacagtgagg 1440
Db 742 atgattcgttcgggtctccgacggggaacgttgatctctgccaagggtgacagtgagg 801
QY 1441 cctctgtcacttcgagaacaatatctgtgtgctgataaggaataacaagctgggtctc 1500
Db 802 cncctgtcacttcgagaacaatatctgtgtgctgataaggaataacaagctgggtctc 861
QY 1501 ggcctgtgcacaagcttacacagcagaggtgtacgggaattgattgattcaaggacttg 1560
Db 862 ggcctgtgcacaagcttacacagcagaggtgtacgggaattgattgattcaaggacttg 921
QY 1561 attatcgacaataatgaggagacagcgttaatacagatgcttcgtctcctgaagctgttc 1620
Db 922 attatcgacaataatgaggagacagcgttaatacagatgcttcgtctcctgaagctgttc 981
QY 1621 tacaagaaaacaatgaggagcgttttgcctcccgctgcagattactcttagagatgat 1680
Db 982 tacaagaaaacaatgaggagcgttttgcctcccgctgcagattactcttagagatgat 1041
QY 1681 tcagaggctcaacttcatttatttaacaagtgaact 1715
Db 1042 tcagaggctcaacttcatttatttaacaagtgaact 1076

RESULT 9
AAZ95004 standard; cDNA; 267 BP.
ID AAZ95004
AC AAZ95004;
XX
XX 15-AUG-2000 (first entry)
DT
XX
XX CSG Prol15 EST useful as prostate cancer marker.
DE
XX
XX Prostate cancer; cancer specific gene; CSG; expressed sequence tag;
KW EST; diagnosis; monitoring; staging; imaging; therapy; metastasis;
KM marker; human; Prol15; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200023111-A1.
PN
XX
XX 27-APR-2000.
PD
XX
XX 19-OCT-1999; 99WO-US24331.
PF
XX
XX 19-OCT-1998; 98US-0104737.
PR
XX
XX (DIAD-) DIADEXUS LLC.
PA
XX
XX Salceda S, Recipon H, Caferkey R;
PI
XX
XX WPI; 2000-339531/29.
DR
XX
XX Diagnosing, staging and monitoring the presence and metastases of
PT prostate cancer especially useful for treating prostate cancer
PT comprises measuring changes in cancer specific gene levels
XX
XX
XX Claim 7; Page 57; 74pp; English.
XX
XX The present sequence is that of cancer specific gene (CSG) expressed
CC sequence tag clone 218985SH1 (Prol15); the full-length contig is
CC given in AAZ95005. The CSG was identified in a database search using
CC the data mining Cancer Leads Automatic Search Package (CLASP), which
CC allows the identification of highly expressed organ and cancer
CC specific genes. Overexpression of Prol15 was observed in 3 of 4
CC primary prostate cancer tissues examined, indicative of it being a
CC diagnostic marker for prostate cancer. The invention provides ESTs
CC and full-length contigs for prostate CSGs (see AAZ94998-295017). The
CC CSGs, polypeptides encoded by them, and antibodies that specifically
```

```
CC bind CSG are used in new, claimed methods for detecting, diagnosing,
CC monitoring, staging, imaging and treating prostate cancer. The new
CC methods provide earlier diagnosis for the presence and metastasis
CC of prostate cancer, and can be used to determine if a cancer has
CC metastasized, or to monitor the progress or stage of the disease
CC when it has not metastasized.
XX
XX Sequence 267 BP; 64 A; 97 C; 53 G; 53 T; 0 other:

Query Match 15.4%; Score 267; DB 21; Length 267;
Best Local Similarity 100.0%; Pred. No. 66-66;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 gaatattccagatacctatcattactctgattgctgtgtgataacagcaagatgcttgaac 123
Db 1 gaatattccagatacctatcattactctgattgctgtgtgataacagcaagatgcttgaac 60
QY 124 tcagggtcacaccagcagctatgtgacacttactatgaanaacatgataccaacggaaac 183
Db 61 tcagggtcacaccagcagctatgtgacacttactatgaanaacatgataccaacggaaac 120
QY 184 cccatcccgacacagcccaactgtgttcgccactgtctacagaggtgcatcccgctcagtac 243
Db 121 cccatcccgacacagcccaactgtgttcgccactgtctacagaggtgcatcccgctcagtac 180
QY 244 taccggtcccccgtgcccacagtaagcccgagggttcctgaacgacagcttccaaacccgtc 303
Db 181 taccggtcccccgtgcccacagtaagcccgagggttcctgaacgacagcttccaaacccgtc 240
QY 304 gtctgcacgcagcccaatcccatcc 330
Db 241 gtctgcacgcagcccaatcccatcc 267

RESULT 10
AAC06945
ID AAC06945 standard; cDNA; 272 BP.
XX
XX AAC06945;
AC
XX
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein 5' EST, SEQ ID NO: 11020.
DE
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PF
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST ) GENSET.
PA
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX
XX WPI; 2000-500381/45.
DR
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
XX Claim 1; SEQ ID 11020; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC cDNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
```


[illegible]

RESULT	13
AACT8475	
ID	AACT8475 standard; cDNA: 2412 BP.
XX	
AC	
XX	AACT8475;
DT	
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	
XX	Human PRO382 (UNQ323) nucleotide sequence SEQ ID NO:68.
KW	
XX	Human: secreted protein; transmembrane protein; PRO; EST; cytosolic
KW	expressed sequence tag; detection; cancer; ss.
XX	
OS	Homo sapiens.
XX	

PN WO2000053756-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 18-FEB-2000; 2000WO-US04341.
 XX
 PR 08-MAR-1999; 99WO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Kljavin JL, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Shelton DL, Stewart TA, Tunas D, Williams PM, Wood WI;
 XX
 DR WPI: 2000-611443/58.
 DR P-PSDB: AAB44230.
 XX
 PT Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities -
 XX
 PS Claim 2; Fig 27; 636pp; English.
 XX
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 CC sequence tag) sequences which encode secreted or transmembrane PRO
 CC polypeptides. The PRO polynucleotides and polypeptides have cytosolic
 CC activity. The polynucleotides and polypeptides can be used for detecting
 CC the presence of PRO polypeptides in samples, for linking bioactive
 CC molecules to cells and for modulating biological activities of cells,
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.
 CC The polypeptide pairs provide specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.
 CC
 XX
 Q0 Sequence 2412 BP; 529 A; 648 C; 643 G; 592 T; 0 other;

[illegible]

QY	649	tacggcgcgcgcgctgcgcgagcgacatgagggtatagaataatatttactctcagccaaaga	708
Db	534	tacggaataatgtctccgtgtgcccaactgggggttcccaagctatagtgagtccagataaactc	593
QY	709	ataagctg_gatgacagcgagatcccaaccagcttatagaaactgaaacaaagctgcgcgcaatgt	767
Db	594	agagctggagctctgcgcgagggcgagcttcgcggagagattgtgtccatccatgcatacctctg	653
QY	768	cgatatctataaaaaactgtaccac-----agtgatgcctgttcttca	810
Db	654	ccgaatbtacaaagvgagactgtcatccacacactcagataatgtatgagggagatgtgtccctc	713
QY	811	aaagcagatggtcttttaagctgtatagcctctcggggtgcataacttgaactgaacgcgcgacag	870
Db	714	ggccagctgtgttaactctgcagctgcacaagcctgtgtgtcatatagaaggggtccacagctccac--	771
QY	871	agcagagatgtgtgagcgagcgagagcgctgcaccgggggacctgtgacccctgtgcagctcagcctg	930
Db	772	----gcatcvggggttggaataacatgcttgccttgctctcgtcagtggtcctgtgcagccagcctc	827
QY	931	caagctccagaacgtccacgcgtgtcggaagctccatcatatatacccccgagtgatcgttgaca	990
Db	828	caattccagaggtctacacccctgcgtggggggtcctgtcatatcagccctctgtgatactcct	887
QY	991	ggccgccacatgctgtggaanaaactcttaacaatcacatgcatgtgacgacatttgcgggg	1050
Db	888	gctgcacacatggtgttatga--ctgttaactccccaagfcatagagc---catccaggtgg	942
QY	1051	attttgagacaactcttcattcgttctctatgagccggatatacgaatagaanaagtattctt	1110
Db	943	gtctagattccctctgttggaataacca--gccccaaccacattgtgtgaaagaattgtctac	1001
QY	1111	catccaattatgtactccacagagacaaagaacaaatgataattgctgtatagaactgcagag	1170
Db	1002	cacagcagaatcacaaagccaaagaaggtcgggcataatgatactgcctctatgaagctgtgcggg	1061
QY	1171	ccctctgaccttcaacgaaactagtgaacacagtgctgtctgcccacccagagatgatactg	1230
Db	1062	ccactcactgcttcaatgaaatgatactcagcctgtgtgcctgcccacactctgtgaagagatctc	1121
QY	1231	cagccagaaacgcgcgtgtgattctccgggttgagggggcacacggagagagaagaagacc	1280
Db	1122	cccgatgtgaaaagatgctgcgtgcagatcagatagaatgtggggccacaagatgtgaggtgtacgcc	1181
QY	1291	tcaagaagtgtctgaacgctgtgccaaagtgctctccatctatgaaagacaaagatgtcaacacgaga	1350
Db	1182	tcacctgtctcgtgaaccacgacgagcgtctccctttgatttccaanaagatctgtcaacacacgg	1241
QY	1351	tatgtctatgtacaacctgtatcacacaacgcatagtactgtgcgcgtcttcttcagaggaaac	1410
Db	1242	gacgtgtgatacgtgtgacatcatctccctccatcatctctctgcgggtgtactctgacaggtgtgc	1301
QY	1411	gtcattctcttgcacgggtgtgcagctgtagagggcctctcgtatcactgtgaacaaacatatctcg	1470
Db	1302	gttgaaacgcgtccacggaggggacagcggggggggccctcgtgtgtgtcaagaagaagaagctgtgg	1361
QY	1471	tgtcgtataaggggatacaaacctgtgggtgtctcgtcgtgtcacaagacttaagaaccaagagt	1530
Db	1362	aagtgtatgtgagagcgaacacattgtgcatcgtcgtgtgcagaggtgtgaacaagcctctgggt	1421
QY	1531	taggggaatttgtatgtattcacaagcgtgtgatttatcgacaatatagaagcgacagcctaa	1590
Db	1422	tacacccgtgtcaactctctcctcgtgactgtgataccacgagcagatgtgagagagacttaaa	1481
QY	1591	tcacacatg 1598	
Db	1482	acctggaag 1489	
RESULT 14			
AAA93842			
ID	AAA93842 standard; cDNA; 2413 BP.		
XX			

AC	AAA93842;
XX	
DT	11-JAN-2001 (first entry)
XX	
DE	Tumour associated differentially-expressed gene 12 cDNA sequence.
XX	
KM	Transmembrane serine protease; TADG-12; chromosome 17; vaccination; ss;
KW	tumour associated differentially-expressed gene 12; cytosolic; human;
XX	malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
XX	
OS	Homo sapiens.
PN	WO200052044-A1.
XX	
PD	08-SEP-2000.
XX	
PF	02-MAR-2000; 2000WO-US05612.
XX	
PR	03-MAR-1999; 99US-0261416.
XX	
PA	(UYAR-) UNIV ARKANSAS.
PI	O'Brien TJ, Underwood LJ,
DR	WPI: 2000-533263/48.
XX	P-PSDB: AAB32246.
XX	DNA fragment encoding tumor associated differentially-expressed gene 12
PT	protein used for diagnosing and treating malignant hyperplasia and
PT	cancers including ovarian cancer -
XX	
XX	Claim 2; Figure 4; 118pp; English.
XX	
CC	This invention relates to a novel transmembrane serine protease called
CC	tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is
CC	located on chromosome 17. Sequences AAA93842-A93845 and AAB32246-B32249
CC	represent human TADG-12 cDNA and their corresponding protein sequences.
CC	A splice variant of TADG-12 (TADG-12V) leads to a truncated protein
CC	product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12
CC	exhibits cytostatic activity, and can be used in vaccines and in gene
CC	therapy. TADG-12 nucleotide and protein sequences are used in the
CC	diagnosis of malignant hyperplasia and cancers or the ovary, breast,
CC	lung, colon, prostate and other cancers where TADG-12 is overexpressed.
CC	TADG-12 is particularly used as tumour marker for early disease
CC	diagnoses. TADG12 proteins or fragments can be used to vaccinate an
CC	individual with cancer, suspected of having a cancer or at risk of
CC	getting cancer. Sequences AAA93846-A93853 represent PCR primers used for
CC	amplifying the TADG-12 cDNA sequence, and in the quantitative analysis of
CC	TADG-12 mRNA. AAB32250 represents a peptide fragment of TADG-12, used to
CC	create anti-TADG-12 antibodies. Sequences AAB32251-B32369 represent
CC	TADG-12 peptides which target HLA, and may be used in a vaccine or for
XX	immune stimulation.
XX	
SQ	Sequence 2413 BP; 554 A; 639 C; 636 G; 564 T; 0 other:
<hr/>	
Query Match	13.0%; Score 225.2; DB 21; Length 2413;
Best Local Similarity	54.4%; Pred. No. 1.2e-53;
Matches 624; Conservative	0; Mismatches 488; Indels 36; Gaps 7
<hr/>	
QY	469 tgcagctcctcagtgactcgatcaaaccccttaactggtgatggcgttcacacgc 528
DB	381 tgcgtcatccttctaagtatactgcgtcataactcgatgtagcaggagctcgattgc 440
QY	529 ccgcgaggggaagaagatcggtgtgttcgctctaaggaacaaattacaccttcg 588
DB	441 aaagacgagggaagacgaagtacccgtgtgtccggtgggtgttcagaatgccgtgcccg 500
QY	589 gttctaccatccaagagagctcctgcacccctgtgttcgaagacagctggaacgaac 648
DB	501 ggtgtcacacgcgtc-----tcgtggaagacatggtgttcgagactggagaaggtcac 554
QY	649 tacgggcgggcgcgcctcgaaggacatbggtctataagaataatttactctagccaaga 708

D 555 taagcaaatctgacctgctcccaactgggtctcccaagctatctgagttcagataaacctc 614
Q 709 atagatg-gatgacagcgaggtcccaagctttatgaacagcaacagcgccgacatgt 767
D 615 agagtgagctcgctcgagggcgaggtcccgaggaggtctgtctccatccgaacctcttg 674
Q 768 gatatctctaaanaaccttaccac-----agtatccctgtcttca 810
D 675 ccagatgacaaggtgactgcatctaacccactcagatatactgagggagggatgtgctctc 734
Q 811 aaagcagtggtctcttcaagctgtatagcttcgagggtaacttgaaacctcagcgccag 870
D 735 ggcacagtggttacctctgacgtgacagacctgtgtgtcatagagaagggtcacagctcac-- 792
Q 871 agcagagtggtggcgagggcgagcgctcccgaggcccttgccctgaggttcagacctg 930
D 793 ----gcatgtgtggaggaaacatgctcctgtcctcgtcagtggtgcctggagcgcaacct 848
Q 931 caagtcagaagcgtccacgtgtgctgaggggtccatcatcaccgccgagtgatcgtgaca 990
D 849 cagttccagggctacacacgtgaggggtcctgtcatcagccctgtggtatcatcact 908
Q 991 gccgcacactgctgtaaaaaaccttacaacatccatgagcatgtgacggcatttcgggg 1050
D 909 gctgacacactgtgttatga--ctgtacctccccaagtcatagtac--catccagtggt 963
Q 1051 attttgagcaactctctctatctctatctgagcgaggtacaaagtaaaaaagtgattctc 1110
D 964 gctctgcttccctgtctgagcaatcca-gccccaacctcctgtgtgagaagatgtcttc 1022
Q 1111 catcaaatatgactcccaagaccaaagaacaaatgacatgctgtgaatgaagctgcaaga 1170
D 1023 caagagaagtacaaagcaaaagagctgagcaatgacatgccttataagagctgcgcgg 1082
Q 1171 cctctgaccttcaaggaaccttaagaaacagtgctgtctgcccaccccaagcatgagctg 1230
D 1083 ccaactacgttcaatgaatgaaatgaaatcagctgtgtgcttcccaactctgaagaacttc 1142
Q 1231 cagcagaacacagctcgtcgtggaattccgggtgagggcgacaggaagaagaaggaagac 1290
D 1143 ccggaatgaaaaagtgctgctgagagctgagaaatgaggggccaacagagaatgaaagtacgc 1202
Q 1291 tcaagaagtgcgtgaacgctgccaaggtgctctcatctatgagacacagagatgcaacagaca 1350
D 1203 tcccccgtccttgaaacacagcgcgctccttgaatttcaacaagatctgcaaccacag 1262
Q 1351 tatgtctatgacaacactgataccacagcatgattctgtgcggcttctctgcaggggaac 1410
D 1263 gacgtgtatgacgtgacatctctccctccatgctctgagggctaacctgaggggtgagc 1322
Q 1411 gtcgaattcttcagaggtgacagtgaggggacctctgtactctgaaagacaatatatctg 1470
D 1323 gtagaacagctgcacaggggagacagcgggggccctggtgtgtgtcaagaaggggagctgtgtg 1382
Q 1471 tggctgataaggggatacaagctgagggtctctgctgtgcaaaagcttacaagcagaagt 1530
D 1383 aagttatgtggagcgagccagcttggtcagctgagcgaggtgaaacagccttgagggtg 1442
Q 1531 tccgggaatgatagtatctacagatgatttaccgaatatgagggcagacgagctta 1590
D 1443 taacaccggtgtaacctctctctgactgtgactccagacagatgagagagaactctaaaa 1502
Q 1591 tccacatg 1598
D 1503 acctgaag 1510

XX 28-MAR-2001 (first entry)
DT Human serine protease #4, encoding HCHAK72 cDNA clone.
XX
DE
XX
KW Human; serine protease; osteopathic; immunosuppressive; antiallergic;
KW antiinflammatory; cytosolic; cardiac; neuroprotective; nootropic;
KW neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;
KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
KW systemic lupus erythematosus; male reproductive system disorder;
KW testicular cancer; digestion and food absorption disorder; arrhythmia;
KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
KW behavioral disorder; Tourette's syndrome; acute myelogenous leukemia;
KW cardiovascular disorder; ocular disorder; drug screening; ss.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 201..1184
FT /tag= a
FT /product= "Human serine protease #4 from clone HCHAK72"
XX
XX
XX MO200068247-A2.
XX
XX 16-NOV-2000.
XX
XX 05-MAY-2000; 2000WO-0512207.
XX
XX
XX 07-MAY-1999; 99US-0133239.
XX 20-MAY-1999; 99US-0135163.
XX 03-AUG-1999; 99US-0147005.
XX 09-SEP-1999; 99US-0152935.
XX 01-NOV-1999; 99US-0162979.
XX
XX (HUMA-) HUMAN GENOME SCT INC.
XX
XX Ruben SM, Shi Y, Young PE, Ni J;
PI WPI: 2000-67979/66.
XX DR P-PSDB: AAY72093.
XX
XX New nucleic acid molecules encoding human serine protease polypeptides,
XX useful for diagnosis, prevention and/or treatment of disorders e.g.
XX osteoporosis, lupus erythematosus and Alzheimer's -
XX
XX
XX Claim 1: Page 272: 289pp; English.
XX
XX The present cDNA sequence encodes human serine protease #4 from clone
XX HCHAK72 (ATCC Deposit No: PTA27).
XX It is used in methods for the diagnosis, prevention and treatment of
XX various disorders related to serine protease such as bone formation
XX disorders (osteoporosis), connective tissue disorders (arthritis),
XX autoimmune disorders (systemic lupus erythematosus), wound healing, male
XX reproductive system disorders (testicular cancer), digestion and food
XX absorption disorders (Crohn's disease), neurodegenerative diseases
XX (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
XX proliferative and cancerous conditions (acute myelogenous leukemia),
XX allergic reactions (asthma), cardiovascular disorders (arrhythmia),
XX ocular disorders (glaucoma) and infectious diseases caused by bacteria,
XX viruses, fungi or parasites. It is also useful for screening therapeutic
XX compounds. Serine proteases are used as immunological probes or
XX polymorphic markers for the identification of chromosomes, cells and
XX tissues in biological samples, identification of male contraceptive
XX agents, delivery of compositions to targeted cells expressing a
XX receptor for serine protease, hybridisation probes and molecular weight
XX markers. Serine protease nucleic acids are also useful in gene therapy.
XX
XX Sequence 1394 BP: 301 A; 381 C; 417 G; 295 T; 0 other;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2001, 09:21:33 : Search time 1348.83 Seconds
(without alignments)
12180.230 Million cell updates/sec

Title: US-09-615-285-1

Perfect score: 1738

Sequence: 1 ggcggagcgcgagcgaggy.....ctggcaaaaaaaaaaaaaa

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

```
117: gb_est148:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query Match	Length	DB	ID	Description			
C 1	492	28.3	529	111	AM058537	AM058537 wx23b08.			
C 2	479.4	27.6	814	168	BF675496	BF675496 60213847			
C 3	465	26.8	465	141	BE857738	BE857738 7947f08.			
C 4	446.4	25.7	450	141	AI393270	AI393270 t099a11.x			
C 5	404	23.2	737	145	BF168078	BF168078 61077633			
C 6	401.4	23.1	411	170	BF858623	BF858623 RC5-FT01			
C 7	400.6	23.0	528	144	BF076876	BF076876 226635 M			
C 8	395.2	22.7	417	23	AI673506	AI673506 wF19h04.x			
C 9	393.2	22.6	961	150	BF584569	BF584569 60209827			
C 10	374.4	21.5	648	144	BF102443	BF102443 60175186			
C 11	370.6	21.3	385	19	AI394553	AI394553 QV8a0a5.x			
C 12	366	21.1	497	170	BF857022	BF857022 t93-FT02			
C 13	365.2	21.0	490	136	BE483122	BE483122 169169 B			
C 14	355.8	20.5	788	159	BF786954	BF786954 60211214			
C 15	338.4	19.5	910	175	BG242090	BG242090 60235463			
C 16	334.8	19.5	790	145	BF144549	BF144549 60179019			
C 17	332	19.1	783	174	BG116274	BG116274 60233803			
C 18	323	18.6	1021	145	BF161225	BF161225 60176947			
C 19	322	18.5	353	140	AI261741	AI261741 q231d04.x			
C 20	321	18.5	703	166	BE376601	BE376601 60122643			
C 21	320.4	18.4	523	140	BE370896	BE370896 237560 M			
C 22	318.4	18.3	328	140	BE772947	BE772947 RC1-FT01			
C 23	313.8	18.1	325	140	BE772923	BE772923 RC1-FT01			
C 24	311.6	17.9	565	20	AI467561	AI467561 ve37e12.x			
C 25	307.2	17.7	564	20	AI480993	AI480993 vF86g09.x			
C 26	301.4	17.3	323	140	BE772937	BE772937 RC1-FT01			
C 27	296.6	17.1	693	141	BE916896	BE916896 60166643			
C 28	296.2	17.0	379	140	BE839313	BE839313 RC6-FN01			
C 29	294.4	16.9	308	147	BF333487	BF333487 QV1-CN00			
C 30	282.6	16.3	1019	141	BE916909	BE916909 60166645			
C 31	274.6	15.8	415	149	BF475681	BF475681 naa10c02			
C 32	268	15.4	300	107	AU099721	AU099721 AU099721			
C 33	266	15.3	557	140	BE838930	BE838930 RC2-FN01			
C 34	266	15.3	564	140	BE772472	BE772472 RC2-FT01			
C 35	266	15.3	567	140	BE772471	BE772471 RC2-FT01			
C 36	266	15.3	572	140	BE772476	BE772476 RC2-FT01			
C 37	264.4	15.2	543	140	BE838932	BE838932 RC2-FN01			
C 38	264.4	15.2	552	140	BE772473	BE772473 RC2-FT01			
C 39	264.4	15.2	560	140	BE838935	BE838935 RC2-FN01			
C 40	263.4	15.2	560	140	BE838929	BE838929 RC2-FN01			
C 41	258.8	14.9	555	113	AW211403	AW211403 u080e12.			
C 42	252	14.5	364	144	BF063194	BF063194 7h85c03.			
C 43	239.4	13.8	255	4	AA244195	AA244195 nc06b12.s			
C 44	234	13.5	528	14	AA87020	AA87020 uc81c11.Y			
C 45	232	13.3	579	140	BE838943	BE838943 RC2-FN01			
ALIGNMENTS									
RESULT 1									
AM058537/c									
LOCUS	AM058537	529 bp	mRNA	EST	09-MAR-2000				
DEFINITION	wx23b08.x1 NCI_CGAP_K1d11 Homo sapiens CDNA clone IMAGE:2544471 3'								
	similar to SW:TM62_HUMAN O15393 TRANSMEMBRANE PROTEASE, SERINE 2 ;								
ACCESSION	AM058537								
VERSION	AM058537.1								
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
	Euteleostomi; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
REFERENCE	1 (bases 1 to 529)								
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .								
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),								
	Tumor Gene Index								

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
www-bio.ljlln1.gov/bdnp/Image/Image.html
Insert Length: 543 Std Error: 0.00
Seq primer: -40UP from GlbCo
High quality sequence stop: 319.

FEATURES
SOURCE
Location/Qualifiers
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:254471"
/clone_1lb="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7MD-Pac (Pharmacia) with a modified polylinker; Site.1: Not T; Site.2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones IDs 1322376-1323911, 1456007-1456775, and 1500532-1502853). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 114 a 150 c 121 g 143 t 1 others

ORIGIN

Query Match 28.3%; Score 492; DB 111; Length 529;
Best Local Similarity 96.8%; Pred. No. 2,7e+131;
Matches 512; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

OY 1084 ggaatccaaagtagaaaagtattcttcataacttaagaaccagaacaacat 1143
Db 529 GGAGCCCAAGTACAAAAGTAGTTCTTCATTCGAATTATGACTGGAGACCAGACAT 470

OY 1144 gaacattgcgtgatgaaagctcgaaagcccttgacctcaacagactagtgaaaccagt 1203
Db 469 GACATGACGCTGTAGTAAMNTCAGAAAGCCTTGACTTTCAACGACTTGTAGTAAACAAGT 410

OY 1204 tgcttggcccaaaccaaggatatcttgtgaagccaagaagcctcttgattccgggtgg 1263
Db 409 TGCTGCCCAACCCAGGCATATGTGTGCAGCCAGATCAAGCTCTGTGTGATTTCCGGGTGG 350

OY 1264 ggagccaccagagaaagggaaagcctcaagaagtgtcgaaagctgcccagaagtgtcttc 1323
Db 349 GGGGCGACCGAGGAGAAGGGAAGACCTCAGAAAGTGTGAACGGTCCCAAGTGCTTTC 290

OY 1324 attgaagaacagagatgcacaagcagatatgtcttatgaacaactgataccaccagcatg 1383
Db 289 ATTGAGACACAGAGATGTGCACAGCAGATATGTCTATGCAACCTGATCACACAGCCATG 230

OY 1384 atctgtgcgcgtcttccttcaggggaaagctgattcttcagaggtgacagtggaagggcct 1443
Db 229 ATCTGAGCCGGCTTCCTCAGAGGGAACCTCCATTCTTCCAGGGGTGACAGTGGAGGCT 170

OY 1444 ctgtgacttcgaaagaacatatctgtgtgctgtgataagggaatacaagcttggg -gtctgg 1502
Db 169 CTGGTCACTTGAAGAACATAATATCTGGGGCTGTGATAGGGGATACCACTGGAGAGTTTGG 110

OY 1503 ctgtggcaaatcttaagcaccaaggagvtlaagggaatgtgltgtattcaaggagctgagt 1562
Db 109 CTGTGCAAAACTTACACACAGGAGGTATAGGGGAATGTGATTCACAGGAGCTGAT 50

OY 1563 ttatcgaaanaatgagggcagagcctaataccacatgtgtcttcgtctctg 1611

circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 9856508-9865759, 1101192-1101959, and 1217928-1220615) subtraction by Dorio Gargese and M. Estime-Denis.

BASE COUNT	101 a	138 c	110 g	116 t
ORIGIN				

Query Match	26.8%;	Score 465;	DB 141;	Length 465;
Best Local Similarity	100.0%;	Pred. No. 1.7e-123;		
Matches 465;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY 1147 atgcgctgaatgaagcgtgcagaagccctgaacttcaacgaacctagtgaaccagttgtgt 1206
Db 465 ATTGCGCTGATGAAGCTGCAGAAGCCTTGACTTCAACGACCTAGTGAACACAGTGTGT 406

QY 1207 cgcgcccaaccagcgtatgctgtcgacccgaacacgctctgtcgattccgggtggyg 1266
|||||
Db 405 CTGCCCAACCCAGGCATGATGCTCGACGCCGAAACACTCTGTGGATTCCGGGTGGGG 346

QY 1267 gccaccggaggaagaaggaaacccagaaagtgcgtgaaagctgtccaagtgtccttcatt 1326
|||||
|||
Db 345 GCCACCGAGGAGAAAGGGAACCTCAGAAGTGTCAACGTCGCCAAGGTGCTTCTCATTT 286

```

qy      1327 gagacacagagatgcaacagcagatatgtctcatgacaaccgatccacaccagcatgac 1386
        |||||
db       285 GAGACACAGATGCCAACGACAGATATGTCTATGACAACCCTGATCAGACCCAGCATGATC 226

```

Db 225 TGTGCGGGTTCCTGCAGSGGACCTGATTCTTGCCAGGATGACACTGGAGGCCCTCTG 166

1447 g tcaatcccgaaagaccacaaatcctcggcggcgcgaataggggaatcaacagcctggggcctcgt 1506
1448 |||||
1449 |||||
1450 |||||
Db 165 gtcacctggaagaacaatattctggtggtcatataggcgatacaagcttgggggtcttcggcctgt 106

Db 105 GCCAAGCTTACAGACCAAGAGTGTACGGATGTGATTTCCACGACGATGATTAT 46

Db 45 CGACAAATGAGGGCAGACGCTAATCCACATGGTCTTCCTTG 1

RESULT	4			
AI393270/c				
LOCUS	AI393270	450 bp	mrna	EST
				30-MAR-1999

ACCESSION	VERSION	REVISION	DESCRIPTION
AI393270	AI393270.1	GI:4222817	similar to TR:O15393 O15393 SERINE PROTEINASE. ;, mRNA sequence.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Carnivora; Canidae; Canis; Canis lupus

REFERENCE
1 (bases 1 to 450)
AUTHORS
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
Tumor Gene Index

JOURNAL COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever.

M.D.,¹ LOUIS M. STAUDY, M.D.,¹ Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

found through the I.M.A.G.E. Consortium/LNL at:

www-bio.lnl.gov/bbrp/image/image.htm
Insert Length: 385 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 309.

FEATURES	Location/Qualifiers
source	1. .450

```

/organism="Homo sapiens"
/lab_xref="taxon:9606"
/clone="IMAGE:210827"
/clone_lib="NCI_CGAP CLL1"
/tissue_type="B-cell, chronic lymphocytic leukemia
/lab_host="DH10B"

```

/note=Vector: pT7SD-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGTGAAGGCGGGCGGCCTTTTTTTTTTTTTTTT T 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7S vector. Library is normalized, and was constructed by Bento Soares and M.Patima Bonaldo."

BASE COUNT	98 a	132 c	107 g	111 t	2 others
ORIGIN					

Query Match	25.7%	Score 446.4	DB 19	Length 450
Best Local Similarity	99.3%	Pred. No. 4.7e-118		
Matches 447; Conservative	0	Mismatches 3	Indels 0	Gaps 0

QY 1162 cggaaagagccctgacttctcaagacatfagaacacagfctcggcccaaccagc 1221
Db 450 ctgcagaaagccctgacttctcancgacttagcgaacacagtgctcggcccaaccagc 391

QY	1222	390	Db
atgatgctgcacccagagacagctcgttgattccggctgggggagccacccagagagaa	1281	atgatgctgcacccagagacagctcgttgattccggctgggggagccacccagagagaaa	331

[illegible]

1342 AACAGCGAGTATGCTCTATGACCAACCTGATCAACCAAGCCATGATCTGTCGGGCTTCCTG 1401
 270 AACAGCGAGTATGCTCTATGACCAACCTGATCAACCAAGCCATGATCTGTCGGGCTTCCTG 211

Db 210 CAGGGGACGTCGATCTTGCCAGGGGTGAGAGGCCCTCTGGTCACCTCGAAGAAC 151

150 AATATCTGGTGGCTGATAGGGGATACAACTGGGGGTTCTGGCTGTGGCCAAAGCTTACAGA 91

90 CCAGGAGTGTACGGGAGTGTGATGGTATTCACGGACGTGATTATCGACCAATGAGG6CA 31

Db 30 GACGGCTATCCACATGCTTCGTCCTG 1

LOCUS	737 bp	EST	30-OCT-2000
BF168078		mrna	
BF168078			
RESULT 5			
BF168078			

mRNA sequence.
 ACCESSION BF168078
 VERSION BF168078.1 GI:11048430
 VERBODEN ECT

SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Db	Accession	Source	Organism	Reference	Title	Journal	Comment
Db	2	TCGATCTCTATATAAAACTCTACCAACACCAATGTTTGTCTTCAAAACATGCTTTCTT					
Qy	827	taagctgtatagccttgcggggatcaacttgaactaagccgcagagcagaatgttggcg					
Db	62	TACCGTGTATAGAAATGCGGGGCTCTCGGTGAAGACAGACCCGCCAAGCCGGATCGGGGCG					
Qy	887	gcgagagcgcgctcccgaggggcgctggcccttgcaggttcagcttcacgttccagaacgctc					
Db	122	GATGCAATGCTTACTCGGCGCAATG6CCTCTG6CAGGTtAGCCTGCACTGCAAGG6CATTC					
Qy	947	acgtgtgtgcgagagctccatccatccaccccgagttgagtcgtltagacgcgcgccatctgtg					
Db	182	ACGTCGTGGGGGCTCCATCATCAACCCCGAGGTGGATGTGACAGCCGCGCATGCGGTAG					
Qy	1007	aaaaaccttcaacaatccatgcatgtggacgtgacgcatlctggcggaattttgagaacatctt					
Db	242	AGGAACCCCTTAACAAATCCCAAGATCTGGGGGCTTTGACAGGAATCTTGAACAAATCTT					
Qy	1067	tcatgttctatgagccgagacacacaagaagagaaaagtatttccatccaatcatatgact					
Db	302	ACATGTTCTATGGAAGTGTGATACCGAGTAGAGCAAAAGTATTTCCACCCAAATATATGTT					
Qy	1127	ccaagaccagaacacatgacatctgcgtgtatgaaagctgtcagaagcctctgacttcaag					
Db	362	CCAAGACCAACAAACAAATGACATTCGCTTAAGAAAGCTCAACACCCCTCTGACTTTCAACG					
Qy	1187	acctagtgaaacacagtggtgtctgcccacccacagacatgcttcacagcagaacagcctt					
Db	422	ACAAAGTGAACCCAGTGCTGTCTTACCCATTCACAGGAGTATCTGAGAGCACTCACTGCTCT					
Qy	1247	gcttgatctccgggttgggggcccacgcgagagaaaggaaagcactca					
Db	482	GCTGGATTTCCGGGTGGGGGCCACCTACCTACGAGAAAGGAGAACCTCA					
RESULT	8						
LOCUS	A1673506/c						
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
COMMENT							
FEATURES							
source							


```

RESULT 10
LOCUS      BF102443      648 bp      mRNA      EST      19-OCT-2000
DEFINITION 601751865F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3979313 5',
            mRNA sequence.
ACCESSION  BF102443
VERSION     BF102443.1
KEYWORDS   GI:10884969
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 648)
            NIH-MGC http://mgs.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE      Unpublished (1999)
AUTHORS    Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
JOURNAL    Tissue Procurement: Gilbert Smith, Ph.D.
            CDNA Library Preparation: Life Technologies, Inc.
COMMENT    CDNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM9173 row: b column: 18
            High quality sequence stop: 633.
FEATURES
    source
        1..648
            location/Qualifiers
            /organism="Mus musculus"
            /strain="FVB/N"
            /db_xref="taxon:10090"
            /clone_image:3979313"
            /clone_lib="NCI_CGAP_Mam1"
            /tissue_type="tumor, biopsy sample"
            /dev_stage="3 months, virgin"
            /lab_host="DH10B"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; Salt:
            Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
            Library constructed by Life Technologies. Investigator
            providing samples: Gilbert Smith, NIH"
BASE COUNT  169 a 154 c 177 g 146 t 2 others
ORIGIN
Query Match      21.5%; Score 374.4; DB 144; Length 648;
Best Local Similarity 78.2%; Pred. No. 3.4e-97;
Matches 499; Conservative 0; Mismatches 133; Indels 6; Gaps 4;
QY 504 ctggtgtgtatggtgtctacacgtcccggtgggagagacgaatcgtgtgttgcct 563
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 3  CTGGTGTGAGGGGGTGAACA-TGTCCCAAGGAGAGATGAGAA-CCGCTGTGTGCTT 60
QY 564 ctacgacccaactcatcctcagggtgtactatctcagaggaagtcctgcaccctgt 623
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 61 CTACGGACAAGGTTTATCTCTCCAGGTTTACTATCTCAGAGGAAGCCGTGATCCCT 120
QY 624 gtgcacaagcagctggaagagaactacggggcggtcctgcaggagacatgggtctataa 663
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 121 gtgcggcggtgtgtgagtgagagctacgagagacgacgtgtaaaacacatgcgatatCAA 180
QY 684 gataatcttctcagcagaagaatagtgatgacagcgatccacacagcttataa 743
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 181 GAACATTTTATTTCTAGCCAAAGGATACCAAGCCAGAGGGGGCAACGAGCTTTATGAA 240
QY 744 actgacacaacatgcccgaatgctgatactataaaaactgtaccacagtgatgcct 803
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 241 GGTGATGTGAGTGTGAGCAATGTGACTCTATATAAACTCTACACACAGTACATCAG 300
QY 804 ttcttcaaaagcagtggttcttcttaccgtgtatagcctcggtggtcaacttgaactcaag 863
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 301 TTTCATCCCGCATGTGTGTTCTTTCGCTGTATAGAAATGCGGGGTTCGCT---CAGTGAA 357

```

```

QY 864 ccgccaagcagagatgtggtgcgagagcggtcccggtgggagccttggtgcaggt 923
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 358 ACCGCCAGCAGAGATTGTGGGTGATTGAATGCTCTACCGAGGAACTGGCCCTGGCAGGT 417
QY 924 cagcctgacgctcagaagctccacgtgtgtcggagaggtccatcatcaccccgagtgat 983
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 418 CAGCTGTCACTCCAAAGCGTCCACGCTCTGCGGAGGGCTCATCATCAACCCCGAGTGTGAT 477
QY 984 cgtgacagccgcccactcgtgtgaaaaaccttacaatcatcatgcatgacgcgcat 1043
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 478 TGTGACGNCGG-CCACTGTGTGGAAGAACCCCTCAGCGGNCACAGGTACTGACGCGCAT 536
QY 1044 tgcgggattttgagacaacttcatgtctctatgtggccggatgtaaaagt 1103
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 537 TGGGGGATTCTGTGAGACAGTCTCATGTTCTATGGAAGTAGACACCGAGGTAGAAAAAGT 596
QY 1104 gattctcatccaatatatgattcccaagaccaaagaaca 1141
    |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  ||
Db 597 AATTTCATCCAAATTAGCAGCTCTAAGACCAAGAAAT 634

```

```

RESULT 11
LOCUS      AI394553/C 385 bp      mRNA      EST      04-FEB-1999
DEFINITION t908405.x1 NCI_CGAP CLL1 Homo sapiens cDNA clone IMAGE:2108144 3',
            similar to TR:O15393 O15393 SERINE PROTEASE. ;, mRNA sequence.
ACCESSION  AI394553
VERSION     AI394553.1
KEYWORDS   GI:4224100
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 385)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
TITLE      Unpublished (1997)
AUTHORS    Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
JOURNAL    Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
COMMENT    CDNA Library Preparation: M. Bento Soares, Ph.D.
            CDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/dbp/image/image.html

```

```

FEATURES
    source
        1..385
            location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_image:2108144"
            /clone_lib="NCI_CGAP_CLL1"
            /tissue_type="B-cell, chronic lymphocytic leukemia"
            /lab_host="DH10B"
            /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker. Site:1: Not I; Site:2: Eco RI; 1st strand cDNA
            was primed with a Not I...oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGAGCGCGGCAATGCTTTTATTTTATTTTATTTT
            T 3']; double-stranded cDNA was ligated to Eco RI
            adaptors (Pharmacia), digested with Not I and cloned into
            the Not I and Eco RI sites of the modified pT73 vector.
            Library is normalized, and was constructed by Bento
            Soares and M.Felima Bonaldo."
BASE COUNT  82 a 116 c 85 g 102 t
ORIGIN

```

```

82 a 116 c 85 g 102 t

```

```

Query Match          21.3%  Score 370.6; DB 19; Length 385;
Best Local Similarity 97.7%; Pred. No. 3.6e-96;
Matches 376; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 1201 ggtgtctgcgccaaccagcagatgctgcgcagccagaacagctctgctgattccggg 1260
Db 385 ggtgtctgcgccaaccagcagatgctgcgcagccagaacagctctgctgattccggg 326
Oy 1261 tggggggccacagagagagaaggaagacccagagatgctgaacgctgcgaaggtgctt 1320
Db 325 ttggggggccacagagagagaaggaagacccagagatgctgaacgctgcgaaggtgctt 266
Oy 1321 ctcaatgagacacagagatgcaacagcagatattctatgacaacctgatacaccaagcc 1380
Db 265 ctcaatgagacacagagatgcaacagcagatattctatgacaacctgatacaccaagcc 206
Oy 1381 atgacctgtgcgggtctctctgcagggggaacgctcgaattcttgcacaggggtgacagtgaggg 1440
Db 205 atgacctgtgtccgggtctctctgcagggggaacgctcgaattcttgcacaggggtgacagtgaggg 146
Oy 1441 cctctgctactcgaagacacatactctgtgctgatatagggatatacaagctcgaggtctt 1500
Db 145 cctctgctactcgaagacacatactctgtgctgatatagggatatacaagctcgaggtctt 86
Oy 1501 ggcgtgccaagctctacagaccagagtgctacggggaatgctgattcaccaggaactgg 1560
Db 85 ggcgtgccaagctctacagaccagagtgctacggggaatgctgattcaccaggaactgg 26
Oy 1561 attatgacaataatgaggcgagcg 1585
Db 25 attatgacaataatgaggcgagcg 1

RESULT 12
LOCUS      BF857022      497 bp      mRNA      EST      16-JAN-2001
DEFINITION QV3-FT0210-031100-436-402 FT0210 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF857022
VERSION     BF857022.1 GI:12244766
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 497)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL    20202663
MEDLINE
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            project. The entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?cli-QV3&ct2-QV3-FT0210-
            031100-436-402&t3=2000-11-03&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 7
            High quality sequence stop: 497.
            Location/Qualifiers
            1..497
FEATURES
source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="FT0210"
/seq_stage="Adult"
/notes="Organ: prostate,tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT      104 a      130 c      137 g      126 t
ORIGIN

Query Match          21.1%  Score 366; DB 170; Length 497;
Best Local Similarity 98.4%; Pred. No. 8.5e-95;
Matches 380; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Oy 521 cacactgccccgcggcgaggagacgaatcggtgtgttcgcctcctacggaaccaactca 580
Db 403 ccccgcccccgcggcgaggagacgaatcggtgttcgcctcctacggaaccaactca 344
Oy 581 tccctcagtgtaacctcattcagaggaagtcctgacccctgtgtgccaaagcagctgga 640
Db 343 ttccctcagtgtaacctcattcagaggaagtcctgacccctgtgtgccaaagcagctgga 284
Oy 641 acgagaactcagggcgggcgctcgaaggacatggtgctataagaataatttactcta 700
Db 283 acgagaactcagggcgggcgctcgaaggacatggtgctataagaataatttactcta 224
Oy 701 gccaaaggaatgtagatgacagcagatccacagccttattgaactgaacacaagtgccg 760
Db 223 gccaaaggaatgtagatgacagcagatccacagccttattgaactgaacacaagtgccg 164
Oy 761 gcaatgctgatatctataaaacctgtaccacagatgagtcgtcttccaagaagcgg 820
Db 163 gcaatgctgatatctataaaacctgtaccacagatgagtcgtcttccaagaagcgg 104
Oy 821 ttcttcaagctgtatagcctgcggggtcaactgaactcaagccgacagacagga-tt 879
Db 103 ttcttcaagctgtatagcctgcggggtcaactgaactcaagccgacagacagga-tt 44
Oy 880 gtgggcgcgcagagacgcgcgtcccgag 905
Db 43 gtgggcgcgcagagacgcgcgtcccgag 18

RESULT 13
LOCUS      BE483122      490 bp      mRNA      EST      28-AUG-2000
DEFINITION BE483122 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  BE483122
VERSION     BE483122.1 GI:9602655
KEYWORDS   EST.
SOURCE      cow.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 490)
AUTHORS    Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
            Wells,K.D.
            Mapping of Expressed Sequence Tags from a normalized bovine mammary
            gland cDNA library
            unpublished (2000)
JOURNAL    Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8414
            Email: tads@psi.barc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred

```

THIS PAGE BLANK (USPTO)

Dh 301 agaaagcactgtgcatcaacccttgagcccttgaggacccctctcgttgaggactgcgtgcg 360
Oy 416 cggccactctggaagtctcatgaggagcaagtgtctcaactctggagacagatgagact 475
Dh 361 cgggctactctcggaaagtctcatgaggagcaagtgtctcaactctggagacagatgagact 420
Oy 476 cctcaagtaactctgcatcaaacccctctaactgtgtgagtggtgtgtcaactgtcccgag 535
Dh 421 cctcaggtaactctgcatcaaacccctctaactgtgtgagtggtgtgtcaactgtcccgag 480
Oy 536 gggagagagcaagaatcgt 595
Dh 481 gggagagagcaagaatcgt 540
Oy 596 catctcag 655
Dh 541 catctcag 600
Oy 656 gggcggcctgcag 715
Dh 601 gggcggcctgcag 660
Oy 716 atgacagcggagatccac 775
Dh 661 atgacagcggagatccac 720
Oy 776 ataaaaaacctgtaac 835
Dh 721 ataaaaaacctgtaac 780
Oy 836 tagctctgcgggttcaacttgaactcaagccgcagagacagatgtgtgtgtgtgtgtgtgtgt 895
Dh 781 tagctctgcgggttcaacttgaactcaagccgcagagacagatgtgtgtgtgtgtgtgtgtgt 840
Oy 896 cgtcccgaggcctgcggcctgcggcctgcggcctgcggcctgcggcctgcggcctgcggcctgc 955
Dh 841 cgtcccgaggcctgcggcctgcggcctgcggcctgcggcctgcggcctgcggcctgcggcctgc 900
Oy 956 gaggctccatcaac 1015
Dh 901 gaggctccatcaac 960
Oy 1016 ttaacaatccatgagcatctggaacgcatctgtgcgggatttttgagacaaatcttcaatgttct 1075
Dh 961 ttaacaatccatgagcatctggaacgcatctgtgcgggatttttgagacaaatcttcaatgttct 1020
Oy 1076 atgagagccgagatcaac 1135
Dh 1021 atgagagccgagatcaac 1080
Oy 1136 agaaacatgacatctgcgtctgataagctgtgcagaaagcctctgaacttcaacagacatgata 1195
Dh 1081 agaaacatgacatctgcgtctgataagctgtgcagaaagcctctgaacttcaacagacatgata 1140
Oy 1196 aaccagtggtctgtcccaac 1255
Dh 1141 aaccagtggtctgtcccaac 1200
Oy 1256 ccgggttgaggggccac 1315
Dh 1201 ccgggttgaggggccac 1260
Oy 1316 tgcctctcatctgagac 1375
Dh 1261 tgcctctcatctgagac 1320
Oy 1376 cagccatgatactgtgtgcggctcctgcagaggagagatgatacttgcgcagggtgtgacagtg 1435
Dh 1321 cagccatgatactgtgtgcggctcctgcagaggagagatgatacttgcgcagggtgtgacagtg 1380
Oy 1436 gaggccctctgttcaacttccagaaac 1495
Dh 1381 gaggccctctgttcaacttccagaaac 1440

Oy 1496 gtcttgctgttccaaagcttacaagacagagtgtaaggaatgtgatgtattcaacg 1555
Dh 1441 gtcttgctgttccaaagcttacaagacagagtgtaaggaatgtgatgtattcaacg 1500
Oy 1556 actgattatctgacaaatgtaggagcagagctaatccaatgtcttctgttaagt 1615
Dh 1501 actgattatctgacaaatgtaggagcagagctaatccaatgtcttctgttaagt 1560
Oy 1616 cgtttacaagaagaacaaatgtaggagctgttcttcccggtcatgattactttaag 1675
Dh 1561 cgtttacaagaagaacaaatgtaggagctgttcttcccggtcatgattactttaag 1620
Oy 1676 atgattcaagagctcaacttatttatttaacaagatgtaactgtctgcg 1723
Dh 1621 atgattcaagagctcaacttatttatttaacaagatgtaactgtctgcg 1668

RESULT 2
US-09-342-749-1
; Sequence 1, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Mong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TPRS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; EARLIER FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1476)
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (724)
; OTHER INFORMATION: Listed as T in Genbank Accession NO. U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (985)
; OTHER INFORMATION: Listed as C in Genbank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1347)
; OTHER INFORMATION: Listed as C in Genbank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1466)
; OTHER INFORMATION: Listed as A in Genbank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1471)
; OTHER INFORMATION: Listed as A in Genbank Accession No. 6166194 U75329.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (478)
; OTHER INFORMATION: This base can be G or A with G being the more
; OTHER INFORMATION: common allele. The codon will change from Val to
; FEATURE:
; NAME/KEY: allele
; LOCATION: (777)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. The codon is unaffected with both

Db	721	tgatagcgtgcygggccaactctgaaatccaaagcccgcaagacagatctcgttggcgcgag	780
Qy	892	agcagctcccccggggccctctggccctctgcaagttaagccttcaagcttccaaagctg	951
Db	781	agcgctcccccggggccctctggccctctgcaagttaagccttcaagcttccaaagctg	840
Qy	952	tgcgagagcttccatcatcatcaacccccagctgatactgtacagaccgcgccactgcgtgnaaaa	1011
Db	841	tgcgagagcttccatcatcatcaacccccagagtgatctgtacagaccgcgccactgcgtgnaaaa	900
Qy	1012	ccctttaaacaatccatgatgatcttggacggcatttgcgggagattttagaanaattcttcatg	1071
Db	901	ccctttaaacaatccatgatgatcttggacggcatttgcgggagattttagaanaattcttcatg	960
Qy	1072	ttctatgtagagccgatatacccaagttagaanaaagtatttcttccatccaaattatgatctccag	1131
Db	961	ttctatgtagagccgatatacccaagttagaanaaagtatttcttccatccaaattatgatctccag	1020
Qy	1132	accaagaacaatgatgacattgcgcgtgtatgaagctgtgcagaagccctctgacttccaagccta	1191
Db	1021	accaagaacaatgatgacattgcgcgtgtatgaagctgtgcagaagccctctgacttccaagccta	1080
Qy	1192	gtgaacaccagtgctcttgcgcccaacccagcgatgatctgtcgaagccagaagaagctctgcgtg	1251
Db	1081	gtgaacaccagtgctcttgcgcccaacccagcgatgatctgtcgaagccagaagaagctctgcgtg	1140
Qy	1252	attccgggtggggggccacccggagagaaaggaaagaccttccgaagctgtctgaacgtctcc	1311
Db	1141	attccgggtggggggccacccggagagaaaggaaagaccttccgaagctgtctgaacgtctcc	1200
Qy	1312	aaggtgcttcttcatcttgagacacagagatgtaaacacagagatalgtctatgtacaaacctgtac	1371
Db	1201	aaggtgcttcttcatcttgagacacagagatgtaaacacagagatalgtctatgtacaaacctgtac	1260
Qy	1372	aacaccagccatgatctgtctcccgcttccctgtcgaagggaagacgttgatcttgcgaaggtgac	1431
Db	1261	aacaccagccatgatctgtctcccgcttccctgtcgaagggaagacgttgatcttgcgaaggtgac	1320
Qy	1432	agtggaagggccctctggtcactcttgaaagaaacaatatctgtgtgctatagaaggatacaagc	1491
Db	1321	agtggaagggccctctggtcactcttgaaagaaacaatatctgtgtgctatagaaggatacaagc	1380
Qy	1492	tggggcttctgctgtgtgcacaaagcttcaagaccagagtgtaagggaatgtgaatgtatctc	1551
Db	1381	tggggcttctgctgtgtgcacaaagcttcaagaccagagtgtaagggaatgtgaatgtatctc	1440
Qy	1552	acggactggaattatcgacaacaatgagggcagacgctaa	1590
Db	1441	acggactggaattatcgacaacaatgagggcagacgctaa	1479
RESULT 3			
US-08-807-151-2			
: Sequence 2, Application US/08807151			
: Patent No. 6043033			
: GENERAL INFORMATION:			
: APPLICANT: Bandman, Olga			
: APPLICANT: Lal, Preeti			
: TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED			
: TITLE OF INVENTION: PROTEASE			
: NUMBER OF SEQUENCES: 5			
: CORRESPONDENCE ADDRESS:			
: ADDRESSEE: Incyte Pharmaceuticals, Inc.			
: STREET: 3174 Porter Drive			
: CITY: Palo Alto			
: STATE: CA			
: COUNTRY: US			
: ZIP: 94304			
: COMPUTER READABLE FORM:			
: MEDIUM TYPE: Diskette			
: COMPUTER: IBM Compatible			
: OPERATING SYSTEM: DOS			

SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807,151
FILING DATE: Filed Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0227 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1077 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCORNO701
CLONE: 556016
US-08-807-151-2

Query Match 57.1% Score 992.4; DB 3: Length 1077;
Best Local Similarity 99.8% Pred. No. 1.6e-260;
Matches 993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 721 agcgagatccacagcttctgaactgaacacaaagtcgagcaatgctacatactataa 780
DB 83 ACCGATCCACAGCTTATGAAGTGAACACACAGTCCGCAATGTCATATCTATAA 142
QY 781 aacagttaccagaagtgatccgtctctcaaaagcagtgcttcttaagctgaagcc 840
DB 143 AACGTGTACACAGTGTATGCTTCTTCAAAAGCAGTGTCTTTACGCTATAGCC 202
QY 841 tgcggggtcaacttgaaaccaaagccgagcagagatctgtgagcgagagcgagc 900
DB 203 TCGGGGTCAACTTGAACCAAGCCCGACAGAGATGCTGGCGCGCAAGCGCGCTC 262
QY 901 ccgaggggcttgccctgagagtcagcctgcaagtcagagacgltccagtgagggc 960
DB 263 CCGGGGGCTTGCGCTGAGCTCAAGCTCAAGACGTCACAGTGTGCGGAGGC 322
QY 961 tccatcatcaccgccgagtgagtgatgagacgcccactgctgtaaaactcttaac 1020
DB 323 TCCATCATCACCCCGAGTGAATCGTACAGCGCCCACTGCTGGAATAACCTTTAAC 382
QY 1021 aatcgaatgcatctgagcagcatctgcgggagcttctgaacaaactcttcaatgga 1080
DB 383 ATTCATGCGCATTTGGACGCGCATTTGGGGGATTTTGAACATCTTTCAATGTTATGA 442
QY 1081 gccgagatccaaagtgaaaaagtgatcttcaatccaaltatgactccaagaacaaac 1140
DB 443 GCGCGATACCAAGTGAATAAGTGAATTTCTCATCCAAATTAAGACTCCAAGCAAGAAC 502
QY 1141 aatgacattgagctgaatgagtgagagcctctgacttcaagagactaagtaaaaca 1200
DB 503 AATGACATTTGCCCTGATGAAGCTGCAAGAGCCTCTGACTTTCAACACACTAGTAAACA 562
QY 1201 gctgtctgcccacccaagcaatgaatgctgagcagcaagcctctgctggaattccgg 1260
DB 563 GTGTCTGTGCCAACCAACGAGCATGATGCTGCAAGCAACACTCTGCTGATTTCCGGG 622
QY 1261 tggggggccacccgagaggaaggaagaccctcaagaatgctgtaacgctgccaagtgctt 1320
DB 623 TGGGGGGCCACCGAGAGAAAGGAAAGCACTCAGAAAGTCTGAACGCTGCCAAGGTGCTT 682
QY 1321 ctcatgtgagacagaagtgcacaaagcagatattgtctaaagaaactgtatcaacagcc 1380
|||||

DB 683 CTCATTGAGACACAGAGATGCAACAGCAGATATGCTTATGACAACTGTATCACACGCC 742
QY 1381 atgactgtgccgctctccgcaagggaaacgctcgaattcttccaggggtacagtggaag 1440
DB 743 ATGATCTGTGCCGCTCTCTGCAAGGGAACGTCGATCTTGGCAGGGTACAGTGGAGG 802
QY 1441 cctctgttacttcgaagaacaatctgtgtgctgataagggaataaagcctggagctt 1500
DB 803 CMTGTGTCACCTGCAAGCAATATCTGTGGCTGATAGGGGATACAAAGCTGGGGTCT 862
QY 1501 ggcctgtgccaaagcttaagcagagagtgatgggaagtgaatgtattcaagcagcgg 1560
DB 863 GCGTGTGCCAAAGCTTACAGACAGAGAGTGAAGGAATGTATGTATTCACGACCTGG 922
QY 1561 attatcgacaatagaggcagacgctaatcacaatgctctgctccttgagcgtctt 1620
DB 923 ATTTATCGACAAATAGGCGCAACGCTTAATCAGATGCTCTTCCCTTGAACGCTTT 982
QY 1621 tacaagaacaatgaggctgtgttcttcccgctgcatgaattactcttagagatgat 1680
DB 983 TACAAGAAACAATGGGCGCTGTTTGTCTCCCGTGCATGATTAATCTTAGAGATGAT 1042
QY 1681 tcagagtgactctatttatttaaacagtgact 1715
DB 1043 TCAGAGTCACTTATTATTAACAGTGAAT 1077

RESULT 4
US-09-008-271A-18
Sequence 18, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Hillman, Jennifer L.
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:

Db 574 TTATGATGGGACCAACACTCTGTGTGGGGGGTCCCTGCTGTCTGGGGACTGGGTCTGACTTC 633

Qy 993 cgcaccatgcgtctgaaaaaacctcttaacaatccatgagcatctgagcgcaatctgcggagat 1052

Db 634 TGCACATTGGCTTTCCAGAGCCGAGACCGGGTCTGTCTGTCGCGTGGCCAGATATTTGGCTGTC 693

Qy 1053 ttctggacaatcttcca-----tgtctataaggagccggaataccagattaga 1097

Db 694 TCTTACCCCGGACCTCACCCCATGCTGTGCACTGTGGGGGTTTCAGGCTGTGATCTTATCATGG 753

Qy 1098 aaaagtgatcttcataaccaaatlatgactccaaagcaacaaatgatcatctgcgtgat 1157

Db 754 GGGCTACCTTCCCTTTCCGAGACCTACTATCATCAGACAAAACACCAATGATGACTTGGCT 813

Qy 1158 gaagctgcgaagaagcctctgactcttcaaacgaagcctagtgaaacgaagctgctcccaacc 1217

Db 814 CCACCTCTCTACTCTCCCTGCTCTCTACAGAAATATATCCAGCCAGTGTGTCTCTGCTGC 873

Qy 1218 aggcatagtatctgcagcagaacaagctctgctcgtgatalctccggtggtggggccaccgaaga 1277

Db 874 GGGACAGGCGCCGTGGATGGCAGAGTGTGTACTGTGATGACCGGCTGGGGTAAACACAGATT 933

Qy 1278 gaagaggaaagacctcagaagtgctgaaagctgccaagtgctcttcatttgaaacacagag 1337

Db 934 CTATGGCCCAACAGGCTATGTGTGCTCTCAAGAGGCCCGGGTGTCCCATCATTAAGACGAAGT 993

Qy 1338 atgcacaagcagaatatagtctatgacaacactgataccaaacgaacatgatactgtgcgagct 1397

Db 994 TTGCAACAGCCCGCACTTCTTACGGGAAATCATGATCAAGCCCAAGATGTTGTGTGCTGGCTA 1053

Qy 1398 cctgcagagggaagcgtcgatctctccagagggtgacaagctggaaggcctctgct----- 1448

Db 1054 TCTCTAGGGGTGGCAATTGATGATCGTGGCAGGGGCGACAGTGGAGGCCCTTGTGTGTGAAGA 1113

Qy 1449 ---caactcgaagaacaatatctggtgcgtgataaggagatacaagctggtctcgagctg 1505

Db 1114 CAGCATCTCTGGGACATCAAGGTGGCGGCTATGTGGCATTTGTAACCTGGCGTACGGGCTG 1173

Qy 1506 tgcacaagcttcagaccagaaggtgtaagggaattgattgattatccacgactgattta 1565

Db 1174 TGCCTTTGGCCCCGGAAGCCAGAGGTGTATACCACTCAAGTCACTGCTCCGAGCTGGATCTT 1233

Qy 1566 tcgacaatatgagggcagacggtatacca 1594

Db 1234 CAGGCCATTAAGACTCAGCTCCGAAGCCA 1262

RESULT 6

US-09-510-738A-188

: Sequence 188, Application US/09510738A

Patent No. 6268165

GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of

TITLE OF INVENTION: Ovarian Cancer

FILE REFERENCE: D6223CIP-A

CURRENT APPLICATION NUMBER: US/09/510,738A

PRIOR FILING DATE: 2000-02-22

PRIOR APPLICATION NUMBER: 09/039,211

PRIOR FILING DATE: 03-14-1998

NUMBER OF SEQ ID NOS: 188

SEQ ID NO 188

LENGTH: 1783

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: full length cDNA of hepsin

US-09-510-738A-188

Query Match	7.18;	Score 124.2;	DB 4;	Length 1783;
Best Local Similarity	51.38;	Pred. No. 2.3e-24;		
Matches 384;	Conservative 0;	Mismatches 338;	Indels 27;	Gaps 3;

QY	873	cagagatgtaggagcgagagagcgcgctcccgaggcgctcgccctgagcgatctagagctcga	932
Db	728	cgcgaatcgtaggagagccgggagacacagctgtagcgagtagcgagtagaagtagatctcgg	787
QY	933	cgctccagaagccgccagctgtagcgagagctctcatcagacaccccgagtagatctgtagagcg	992
Db	768	ctatgagtagagacaaacctctgtaggtagatccctgctctcccgagtagcgtagctgtagagcg	847
QY	993	cgcgcacatcgtagaanaaacctcttaacaatcctagagcatgtagacgagcatctgcggagat	1052
Db	848	cgcgcacatcgctcccgagagcggaacccggtctcctgctccagtagcgagtagctgtagcgctgc	907
QY	1053	tttagagcaatcttcaatgctctat-----ggagcgcgagataccaagttagaanaaagt	1103
Db	908	cgtggtcccgagcgccctcccgacaggtctgcagcgtgggtgtagagcgtgtagcttaacacag	967
QY	1104	gattctcatcaaatatatactacccaagaacaaag-----caatgagcatgtagctgcat	1157
Db	968	gggtctatcttccctcttcggagaccccaacagcgagaggaacagcaacgatatgtccctgct	1027
QY	1158	gaagatctgcagaagacctctgagcttcaacgaacgaactagtagaaacagtagtctgccaaccc	1217
Db	1028	ccacctctcgaagctccctcgccctcccaagatacatcacagcgtgtagctcccgagctgc	1087
QY	1218	agggatgtagttagagcagcagaacagctctgctgtagtatcttcgggtgtaggggccaacggag	1277
Db	1088	cggcccgagccccctgtaggtagagcaagatctgtagcggtagagggcgtgagggaacacgacga	1147
QY	1278	gaagagggaagacaccccaagaagtagctgtagaacgcgtgcgaagagtagcttccatctgagacaaag	1337
Db	1148	ctatgtagcaaacagcgcggttactccaagaggtctgtagtcccatatcagaatagatgct	1207
QY	1338	atgcaacagcagagatagtctcatatgaacaacctgatacacacgaacgaatgctgtagcgctt	1397
Db	1208	ctgcaatagcgctgtagcttctatgtagaaccgaacgaacgaacgaatgctctgtagcgctc	1267
QY	1398	cctgcagaggaagcgtgcatatctctgcgaagggtagacatgtagagtagagctctcgctg-----	1448
Db	1268	cctccgagggtagcatgtgtagtgcctgcgcgaagggcgacagcgtagtgcctcttgtagtagag	1327
QY	1449	---caattcgaaagaaacaatactctgtagtagatagagggagatacaagctgtaggtctctgagctg	1505
Db	1328	cagcatctctcgagcgacagcaggttcgtagcgctgtagcatgtagtagtagtagtagtagtagtag	1387
QY	1506	tgcgaagcttacaagcacaagagtagtagcggaatgtagtagtatctcaacgaacgtgattta	1555
Db	1388	tgcctctgcccagagagcacaagcgctctacacacaaagtagtagctctccgagtagtagatctt	1447
QY	1566	tcgacaatagagggcgagcgcttaacga	1594
Db	1448	ccagggccataagctctacatccgaagcgca	1476

```

1  RESULT 7
2  US-08-200-900A-1
3  ; Sequence 1, Application US/08200900A
4  ; Patent No. 5663566
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT:
9  ;
10 ; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
11 ;
12 ; NUMBER OF SEQUENCES: 38
13 ;
14 ; CORRESPONDENCE ADDRESS:
15 ;
16 ; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
17 ;
18 ; STREET: 87 Cambridgepark Drive
19 ;
20 ; CITY: Cambridge
21 ;
22 ; STATE: MA
23 ;
24 ; COUNTRY: USA
25 ;
26 ; ZIP: 02140
27 ;
28 ; COMPUTER READABLE FORM:
29 ;
30 ; MEDIUM TYPE: Floppy disk
31 ;
32 ; COMPUTER: IBM PC compatible
33 ;
34 ; OPERATING SYSTEM: PC-DOS/MS-DOS
35 ;

```

```

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200,900A
FILING DATE: 23-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Melnert, Maureen C.
REGISTRATION NUMBER: 31,544
REFERENCE/DOCKET NUMBER: GI 5201-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170 X8574
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2581 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-200-900A-1

```

```

Query Match      6.7%; Score 115.8; DB 1; Length 2581;
Best Local Similarity 50.9%; Pred. No. 5e-22;
Matches 354; Conservative 0; Mismatches 332; Indels 9; Gaps 3;

```

```

QY 874 aggatgtggtggcgagagcgccgctccggggcgctgtgcccgtgagcgtcagcctgcac 933
DB 1688 AAGATTGTGGAGGAGAGTACTCCAGAGAGAGCGCTGGCTGGGTGCTGCTGTGTAT 1747
QY 934 gtccagaacgtccacgtgtgctgagaggtccatcatcaccccccagtgatcgtaagcc 993
DB 1748 TTGAGAGATCAACAGGCTGCGGAGCTTCTCTGTAGCAGGAGATTGGCTGGTGGCC 1807
QY 994 gcccaactgcgtggaacaaaccttcaaatccatggcatgtgagcggcatttgcgggatt 1053
DB 1808 GCCCAGCTGGGTGACGAGAAATATGAGACCGCTTAAGTGGAACACAGTCTAGGCTTG 1867
QY 1054 ttgagcaactctcatg--ttctatgagcgcgatatacaagaagaaaaagtattctc 1111
DB 1868 CATATGGCATCAATGTGACTTCTCTCATGANTGAAACTAGGTGATGACCAAAATGTC 1927
QY 1112 at-----ccaattatgactccaagacaagaacatgacattgcgtgatgaagctgcag 1167
DB 1928 ATAAACCCACACTACATAAAGAGAGAAAGAACATGACATTGCCATGATGATCTTGAA 1987
QY 1168 aagcctctgacttcaacgactagtgaaacagtgctgctgcccaacccagcagatgatg 1227
DB 1988 ATGAAAGTGAACACTACAGATATATATACAGCCATTGTTTACCAAGAAATCAAGTT 2047
QY 1228 ctgcagcagaagaagctctgctgctgattccgggtgagggggcccccgaaggaaggaag 1287
DB 2048 TTTCCCCCGAGGAAGAAATTTGTTCTATTGCTGGCTGGGGGCACTTATATATCAAGTTCT 2107
QY 1288 acctcagaagtgctgaacgctgcgaagtgcttctcatltgagacacagagatcaacagc 1347
DB 2108 ACTGACAGAGTACTGCAAGAGAGTGAGCTTCCCTTCTATCAAAATGAAATGTCA--A 2164
QY 1348 agatatgtctatgacaacactgatacacaagcagatgactgtgctgcggcttctctcaggg 1407
DB 2165 CAACAGATGCCAGAAATATATACATTACGAAAAATATGTTGTCAGGCTATGAAAGCAGGA 2224
QY 1408 aacgtgatctcttgcagaggtgacagtgagggccctctgtaactctcgaagaacaatc 1467
DB 2225 GGGGTGATTTCTTGTCAAGGGGATTCAGCGGACCACTATGTGCCAAGAAACAAACAGA 2284
QY 1468 tggtagctatagaggaatacaagcttgagggttctgctgtgcgaagcttaccagacagga 1527
DB 2285 TGGCTCTGCTGGCTGGCTGACGTCATTTGGATATCAATGTGCACTGCTTATGCCCCAGGG 2344
QY 1528 ggttagcgaagatgtgattatcagagactggat 1562
DB 2345 GTGTATGCCCCGGGTCCCAAGGTTTACAGAGTGTGAT 2379

```

```

RESULT 8
PCT-US94-00616-1
; Sequence 1, Application PCT/US9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US94-00616-1

```

```

Query Match      6.7%; Score 115.8; DB 5; Length 2581;
Best Local Similarity 50.9%; Pred. No. 5e-22;
Matches 354; Conservative 0; Mismatches 332; Indels 9; Gaps 3;

```

```

QY 874 aggatgtggtggcgagagcgccgctccggggcgctgtgcccgtgagcgtcagcctgcac 933
DB 1688 AAGATTGTGGAGGAGAGTACTCCAGAGAGAGCGCTGGCTGGGTGCTGCTGTGTAT 1747
QY 934 gtccagaacgtccacgtgtgctgagaggtccatcatcaccccccagtgatcgtaagcc 993
DB 1748 TTGAGAGATCAACAGGCTGCGGAGCTTCTCTGTAGCAGGAGATTGGCTGGTGGCC 1807
QY 994 gcccaactgcgtggaacaaaccttcaaatccatggcatgtgagcggcatttgcgggatt 1053
DB 1808 GCCCAGCTGGGTGACGAGAAATATGAGACCGCTTAAGTGGAACACAGTCTAGGCTTG 1867
QY 1054 ttgagcaactctcatg--ttctatgagcgcgatatacaagaagaaaaagtattctc 1111
DB 1868 CATATGGCATCAATGTGACTTCTCTCATGANTGAAACTAGGTGATGACCAAAATGTC 1927
QY 1112 at-----ccaattatgactccaagacaagaacatgacattgcgtgatgaagctgcag 1167
DB 1928 ATAAACCCACACTACATAAAGAGAGAAAGAACATGACATTGCCATGATGATCTTGAA 1987
QY 1168 aagcctctgacttcaacgactagtgaaacagtgctgctgcccaacccagcagatgatg 1227
DB 1988 ATGAAAGTGAACACTACAGATATATATACAGCCATTGTTTACCAAGAAATCAAGTT 2047
QY 1228 ctgcagcagaagaagctctgctgctgattccgggtgagggggcccccgaaggaaggaag 1287
DB 2048 TTTCCCCCGAGGAAGAAATTTGTTCTATTGCTGGCTGGGGGCACTTATATATCAAGTTCT 2107
QY 1288 acctcagaagtgctgaacgctgcgaagtgcttctcatltgagacacagagatcaacagc 1347
DB 2108 ACTGACAGAGTACTGCAAGAGAGTGAGCTTCCCTTCTATCAAAATGAAATGTCA--A 2164
QY 1348 agatatgtctatgacaacactgatacacaagcagatgactgtgctgcggcttctctcaggg 1407
DB 2165 CAACAGATGCCAGAAATATATACATTACGAAAAATATGTTGTCAGGCTATGAAAGCAGGA 2224
QY 1408 aacgtgatctcttgcagaggtgacagtgagggccctctgtaactctcgaagaacaatc 1467
DB 2225 GGGGTGATTTCTTGTCAAGGGGATTCAGCGGACCACTATGTGCCAAGAAACAAACAGA 2284
QY 1468 tggtagctatagaggaatacaagcttgagggttctgctgtgcgaagcttaccagacagga 1527

```



```

Db 420 AATAAGTAATGATGTAATGTAATGACACCATAGTATTAATGAGCCATCTTGTGGAAAT 479
Qy 1383 gattcgtgcggctccctcgaagggaagtcgactcttcgaagggtgacagtggaagggcc 1442
Db 480 GCTGTGCTGGAGTAACCTCAAGTGTGAGTGTGACGATCTCAAGGCTGACTTGTGTGGCCC 539
Qy 1443 tcttggtactctgaa--gaacaatactggtggtcgtatagggatatacaagcttggtgttc 1499
Db 540 ACTAGTACAAGAAAGATCAAGCGCGCTTGTATTTATTTGTGGGATAGTAAGCTGGGGAGA 599
Qy 1500 tggcgtgtcgaagacttaagaccaggagtgatcaggaatgtagatgtatcagagctg 1559
Db 600 TAGTGTGCTGCGCCGATTAAGCAAGAGTGTATCTCGAGTGACAGCCTTACCTTGACTG 659
Qy 1560 gattatcagcaaatgag 1578
Db 660 GATTAGGCAACAACCTGG 678

RESULT 12
US-07-882-202A-3
; Sequence 3, Application US/07882202A
; Patent No. 5374617
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James H.
; APPLICANT: Comp, Philip C.
; TITLE OF INVENTION: Treatment of Bleeding with Modified
; TITLE OF INVENTION: Tissue Factor in Combination with FVIIa
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richards, Medlock & Andrews
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: Texas
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/882,202A
; FILING DATE: 13-MAY-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hansen, Eugenia S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: OMRF B34290
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214-939-4500
; TELEFAX: 214-939-4600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Blood
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1433
; OTHER INFORMATION: /note= "Coding portion of human
; OTHER INFORMATION: factor VII cDNA"
US-07-882-202A-3
Query Match 5.5%; Score 95.4; DB 1; Length 1440;

```

```

Best Local Similarity 47.9%; Pred. No. 1.4e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;
Qy 865 ccgcagagcagagattgtggtggcgagagcgcgcctccgggggcttggtcccggaagtc 924
Db 660 CCCCAGGCGCAATGTGGGGGCAAGGTGTCCCCAAAGGGAGGTCTCAATGAGCGATC 719
Qy 925 agcttgacgtcccaagacgtccacgtgtgctgagaggtcccatcatcaaccctgagtc 984
Db 720 CTGTTGTTGTAATGAGACTCAGTTGTGTGGGGAGCCCTATCAACCATTCGTGGGTG 779
Qy 985 gtgacaagccgcgaactgtgtgaaacaccttaacaatcatgcatgtgaagcatlt 1044
Db 780 GTCTCGCGCCGCTCCTGTTGACAAATCAAGAACTGAGGAACTGATCCGCTGCTG 839
Qy 1045 gcgggattttgagacaatcttcattgtctatgttagagccgagataccagtagaanaagt 1104
Db 840 GCGGAGCAGACCTTCAGCAGCAGCAGCGGGATGAGCAGAGCGCGGGTGGCGAGGTC 899
Qy 1105 attctcatccaatattatgactccaagaagcaaatgacattgtgcgtgataagctg 1164
Db 900 ATCATCCCGCAGCAGTACGTCCCGGACACCAACAGACATGCGCTGCTCCGCTG 959
Qy 1165 cagaagcctctgacttcaagacactagtgaaccagtggtgtctgtccaa-----c 1215
Db 960 CACACACCGCTGCTCCTACTGACCATGTGTGCTCCCTGCTGCTCCGACAGCATTC 1019
Qy 1216 ccaggatgatgtcgaagcgaagacgtctgtgatttccgggtggtggggcgaagcag 1275
Db 1020 TCTGAGAGAGACGCTGCGCTTCGCTTCATTTGCTGACGGGCTGGGCGCAGCTGCTG 1079
Qy 1276 gagaagaaggaagcctcgaagtgctgaagcgtgcgaagtgcttctcatgtagaacag 1335
Db 1080 GACCGGCGCGCCACGCGCTCGAGCATGTGTCACAAAGCTGCCCGCTGATGACCCAG 1139
Qy 1336 agatgc-----aacgcgatatgtctatgacaactcatcaaccagcatgac 1386
Db 1140 GACTGCTTCACACAGTCAACGAGGTGGAGACTCCCAATATATCAGGAGTACATGTTTC 1199
Qy 1387 tgtccggcttcctgcaagggaagcgcgactcttcgaagggtggaagtgagggcctctg 1446
Db 1200 TGTGCCGCTACTCGATGTGCAAGAGACTCTGCAAGGGGAGAGTGAAGGCCACACT 1259
Qy 1447 gtcaactcgaagaacaatactgtgtgctgataagggatatacaagcttggtgtctgt 1506
Db 1260 GCCACCACTACCGGGGCGACGTGTACTGTAGCGGCACTGCTGCGTGGGCGCAGGCTGC 1319
Qy 1507 gccaaagcttaagaccaggagtgatcagggaaatgtagatgtatcaagagactttat 1566
Db 1320 GCAACCGTGGCCACTTGTGGGTGTACACAGAGGTCTCCCATACATCGATGCTGCNA 1379
Qy 1567 cgaacaatgagggcaga 1583
Db 1380 AAGCTCATCGCTCAGA 1396

RESULT 13
US-08-021-615A-3
; Sequence 3, Application US/08021615A
; Patent No. 5504064
; GENERAL INFORMATION:
; APPLICANT: Morrissey, James H.
; APPLICANT: Comp, Philip C.
; TITLE OF INVENTION: Treatment of Bleeding with Modified
; TITLE OF INVENTION: Tissue Factor in Combination with an Activator of
; TITLE OF INVENTION: FVII
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richards, Medlock & Andrews
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: Texas
; COUNTRY: US

```



```

; NAME/KEY: CDS
; LOCATION: 36..1433
; OTHER INFORMATION: /note="Coding portion of human
; OTHER INFORMATION: factor VII cDNA"
US-08-321-777-3

```

```

Query Match      5.5%; Score 95.4; DB 1; Length 1440;
Best Local Similarity 47.9%; Pred. No. 1.4e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

```

```

QY 865 cgcacagagagatgtgtggcgagagcgctcccgggcccgccctggcagctc 924
DB 660 CCCCAAGGCCGGAATGTGGGGGCAAGGTGCGCCCAAGGGGAGTGTCCATGCGCAGGTC 719
QY 925 agccttcacgtccagaacgtccagctgtcgagaggtccatccaccccgagtgatc 984
DB 720 CTGTTGTGTGTAATGAGACTCAGTTGTGTGGGGGAGCCCTGATCAACACATCTGGGTG 779
QY 985 gtacagagccgcacactgcgtlgaaacacctbaacaatccatgcatctgacgcatctt 1044
DB 780 GTCTCCGCGCGCCACTGTTTCGACAAATCAAGAACTGAGAGAACTGATCGCGGTGCTG 839
QY 1045 gcgggagattttagaacaatcttcatgtctatgtgagccgagataaccagtagaanaagtg 1104
DB 840 GCGGAGCACGACCTCAGCAGCAGCAGGAGATGACAGACCGCGGCGGTGCGCAGGTC 899
QY 1105 attctccatccaaattatgtactccaaagacaaagaaatgacatgtgcctgataagctg 1164
DB 900 ATCATGCCAGCAGCAGTACTCCCGGGCACCACCAACGACATGCGCTGCTCCGCTCG 959
QY 1165 cagaagccttcgacttccagaacagctagtgaacacagtgctgtcccaa-----c 1215
DB 960 CACGACCGCGGCTGCTCCTCACTACCATGTGTGTCCTCTGCTGCTGCCGACGAGCATGTC 1019
QY 1216 ccaagcatgtatgtcagagcagaacagctcgtcgtgattccgggtggggggccacagag 1275
DB 1020 TCTTGAGAGGAGCTGTGCTTCTGCGCTTCTCATTTGTGTAGCGGCTGGGGCCAGCTGTC 1079
QY 1276 gagaagagagaaagcctccagaagtgctgaagcgtcgtccaaagtgctctccatctgagacag 1335
DB 1080 GACCGTGGCGCCACGCGCCCTGGAGTCAATGTGTGCTCAAGTGTGCGCGGCTGATGACCCAG 1139
QY 1336 agatgc-----aacagcagatatgtctatgacaacctgataccaccatgatac 1386
DB 1140 GACTGCTGTCACAGTACAGGAGGTGGAGACTCCCAAAATATCACGAGTACATGTTTC 1199
QY 1387 tgtgcggagctcctcgtcagagagagctcgtatctctgcccaggtgagagagcctctg 1446
DB 1200 TGTGCGGCTACTCGGATGGCAGCAAGACTCTCTGCAAGGGGAGACAGTGGAGGCCACAT 1259
QY 1447 gtacacttcgaagaacaatactctgtgtcgtatagagagatatacagactgggtctctggctg 1506
DB 1260 GCCACCCACTACCGGGGCGCAGTGTACTGTACGAGGGCATGTCAGCTGGGGCCAGGGCTGC 1319
QY 1507 gccaaagcttaacagacagagtgtaagggagtgatgtagtattcaagcagtgatattat 1566
DB 1320 GCAACCGGTGGGCGCACTTGTGGGTGTACACACAGGCTCTCCAGTACATCGATGGGTGCAA 1379
QY 1567 cgaacaatgagggcaga 1583
DB 1380 AAGCTCATGCGCTCAGA 1396

```

```

RESULT 15
US-09-009-217-13
; Sequence 13, Application US/09009217
; Patent No. 6132729
; GENERAL INFORMATION:
; APPLICANT: Thorpe, Philip E.
; APPLICANT: King, Steven W.
; APPLICANT: Gao, Boning
; TITLE OF INVENTION: COMBINED TISSUE FACTOR AND

```

```

; TITLE OF INVENTION: CHEMOTHERAPEUTIC METHODS AND COMPOSITIONS FOR COAGULATION
; TITLE OF INVENTION: AND TUMOR TREATMENT
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009 217
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/042,427
; FILING DATE: 27-MAR-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/036,205
; FILING DATE: 27-JAN-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/035,920
; FILING DATE: 22-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: UTSID:536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-009-217-13

```

```

Query Match      5.5%; Score 95.4; DB 3; Length 1440;
Best Local Similarity 47.9%; Pred. No. 1.4e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

```

```

QY 865 cgcacagagagatgtgtggcgagagcgctcccgggcccgccctggcagctc 924
DB 660 CCCCAAGGCCGGAATGTGGGGGCAAGGTGTCGCCCAAGGGGAGTGTCCATGCGCAGGTC 719
QY 925 agccttcacgtccagaacgtccagctgtcgtgagaggtccatcatcaaccccgagtgatc 984
DB 720 CTGTTGTGTGTAATGAGACTCAGTTGTGTGGGGGAGCCCTGATCAACACATCTGGGTG 779
QY 985 gtacagagccgcacactgcgtlgaaacacctctaaacaatccatgcatctgacgcatctt 1044
DB 780 GTCTCCGCGCGCCACTGTTTCGACAAATCAAGAACTGAGAGAACTGATCGCGGTGCTG 839
QY 1045 gcgggagattttagaacaatcttcatgtctatgtgagccgagataaccagtagaanaagtg 1104
DB 840 GCGGAGCACGACCTCAGCAGCAGCAGGAGATGAGCAGAGCCGGGTGCGCAGGTC 899
QY 1105 attctccatccaaattatgtactccaaagacaaagaaatgacatgtgcctgataagctg 1164
DB 900 ATCATGCCAGCAGCAGTACTCCCGGGCACCACCAACGACATGCGCTGCTCCGCTCG 959
QY 1165 cagaagccttcgacttccagaacagctagtgaacacagtgctgtcccaa-----c 1215
DB 960 CACGACCGCGGCTGCTCCTCACTACCATGTGTGTCCTCTGCTGCTGCCGAGACGAGCTTC 1019
QY 1216 ccaagcatgtatgtcagagcagaacagctcgtcgtgattccgggtggggggccacagag 1275

```

```
Db 1020 TCTGAGAGGAGCGCTGCGCTTCGTGCGCTTCATTGGTCAGCGGCTGGGCCAGCTGCTG 1079
QY 1276 gagaagaaggagacccctcaagaagtctgaacgctgcccagaagtctctcaltgaacacag 1335
Db 1080 GACCGTGGCGCCACGCGCCCTGGAAGCTCATGTGCTCAACGTCGCCCGCTGATGACCCAG 1139
QY 1336 agatgc-----aacaagcagatatgtctatgacaaacctgatacaccagccatgac 1386
Db 1140 GACTGCTGACGACGACGACGAGAGGAGTGGAGACTCCCCAAATATACGAGAGTACATGTTT 1199
QY 1387 tctgcccgtcttcctgcaggggaaacgtcgaatctctgccaagggtgacagtgaggggcctctg 1446
Db 1200 TGTGCGGCTACTCGGATGGCAGCAAGGACTCCTGCAGAGGGGACAGTGGAGGCCACAT 1259
QY 1447 gtcaactcgagaagaacataatctgtgtgctgatataggggataacaagctggggttctgtctgt 1506
Db 1260 GCCACCCACTACCGGGGACAGTGTGACTGTACCGGCATCGTCAGCTGGGGCCAGGGGCTGC 1319
QY 1507 gccaaagcttacagaccagagtgctacggggaatgtgatlgtatcacggaatgatat 1566
Db 1320 GCAACCGTGGGCCACTTGTGGGGGTGTACACCAAGGCTCTCCAGTACATGAGTGGCTGCAA 1379
QY 1567 cgacaaatgagggcaga 1583
Db 1380 AAGCTCATGCGCTCAGA 1396
```

Search completed: September 26, 2001, 09:22:54
Job time: 81 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:05 : Search time 34.76 Seconds
(without alignments)
858.084 Million cell updates/sec

Title: US-09-615-285-2
Perfect score: 2717
Sequence: 1 MALNSGSPPAIGPYENHGY.....YGNVWFETDIYROMRADG 492

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq_0601:*
2: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT:*
3: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT:*
4: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT:*
5: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT:*
6: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT:*
7: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT:*
8: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT:*
9: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT:*
10: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT:*
11: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT:*
12: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT:*
13: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT:*
14: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT:*
15: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT:*
16: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT:*
17: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT:*
18: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT:*
19: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT:*
20: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT:*
21: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT:*
22: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT:*
23: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	492	21	AAV77726
2	2717	100.0	492	21	AAV44406
3	2709	99.7	492	21	AAV92050
4	2696	99.2	492	21	AAV36901
5	2696	99.2	492	21	AAV57280
6	1540	56.7	283	21	AAV81492
7	888	32.7	454	21	AAV32246
8	885	32.6	452	20	AAV41694
9	884	32.5	453	21	AAV44250
10	742.5	27.3	327	21	AAV72093
11	676.5	24.9	435	20	AAV06437

12	676.5	24.9	435	22	AAV72558	Human seripancrin
13	674	24.8	432	21	AAV99417	Human PRO1570 (UNO
14	674	24.8	432	22	AAV87581	Human PRO1570. HO
15	674	24.8	432	22	AAV66166	Protein of the Inv
16	671.5	24.7	432	22	AAV72559	Human seripancrin
17	666.5	24.5	480	21	AAV08950	Human secreted pro
18	665.5	24.5	445	21	AAV11698	Mouse serine prote
19	660	24.3	798	15	AAV57283	Bovine enterokinase
20	651.5	24.0	457	21	AAV11699	Human serine prote
21	649.5	23.9	248	21	AAV43572	Human cancer assoc
22	648.5	23.9	414	21	AAV08912	Human secreted pro
23	601	22.1	1042	21	AAV44426	Human serine prote
24	588	21.6	311	21	AAV11697	Mouse serine prote
25	587.5	21.6	273	21	AAV11696	Mouse serine prote
26	577.5	20.6	1113	21	AAV44427	Mouse Serine prote
27	558.5	20.6	683	21	AAV19551	Human matrilysin
28	558.5	20.6	762	21	AAV90284	Human peptidase, H
29	558.5	20.6	855	21	AAV06671	Tumour antigen der
30	558.5	20.6	855	21	AAV19552	Human matrilysin
31	558.5	20.6	855	22	AAV35465	Human membrane-ty
32	558	20.5	296	21	AAV72108	Human serine prote
33	558	20.5	372	21	AAV72092	Human serine prote
34	544	20.0	238	21	AAV11695	Mouse, serine prote
35	539.5	19.9	416	20	AAV43325	Mouse hepsin prote
36	539.5	19.9	416	20	AAV96812	Human PRO618 prote
37	538.5	19.8	802	20	AAV41710	Human PRO618 (UNO3
38	538.5	19.8	802	21	AAV44266	Human PRO618 (UNO3
39	538.5	19.8	802	21	AAV24052	Human PRO618 prote
40	534	19.7	356	19	AAV46917	Trypsin-like enzym
41	527.5	19.4	418	20	AAV29498	Human lung tumour
42	527.5	19.4	418	20	AAV29501	Human lung tumour
43	527.5	19.4	418	20	AAV29502	Human lung tumour
44	527.5	19.4	418	20	AAV29502	Human lung tumour
45	527.5	19.4	418	21	AAV44437	Human lung tumour

ALIGNMENTS

RESULT 1
AAV77726
ID AAV77726 standard; Protein: 492 AA.
AC AAV77726:
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMPRSS2 polypeptide.
XX
KW Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
KW gene therapy; protein therapy.
XX
OS Homo sapiens.
XX
PN WO20000605-A1.
XX
PD 06-JAN-2000.
XX
PF 29-JUN-1999; 99WO-US14622.
XX
PR 29-JUN-1998; 98US-0091044.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Wong AKC, Tavtigian SV, Teng DHF;
XX
DR WPI: 2000-170914/15.
XX
DR N-PSDB: AAZ87786.
XX
PT Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of
XX human cancer -
XX
PS Claim 55; Page 77-79; 89pp; English.

XX The invention provides a new tumour suppressor gene, designated TMRSS2.
 CC The TMRSS2 polynucleotides and polypeptides can be used in methods for
 CC diagnosing and prognosing predisposition to cancer in humans. The
 CC polypeptides may also be used in assays to screen for compounds with
 CC anti-cancer or therapeutic properties. The polypeptides are also useful
 CC for rational drug design. The TMRSS2 polynucleotides and polypeptides
 CC may be used for gene therapy and protein therapy. The present sequence
 CC represents the TMRSS2 polypeptide.

XX Sequence 492 AA:

Query Match 100.0%; Score 2717; DB 21; Length 492;
 Best Local Similarity 100.0%; Pred. No. 2.9e-203;
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYRPNAPQTYVPTVYEVHNAQYRSPVQYAPRVLTQA 60
 DB 1 malnsghppaigpyenhyqpenypaqrvtyevhpqyypvqyaprvltqga 60
 QY 61 SNPVVCTOPKSPSGVCTSKTKKALCITLTLTGTFVGAALAGLLMKFMGSCNSNGIEC 120
 DB 61 snpvcctopkspsgvctsktkkalcitltltgtflvgaalaagllmkfmgscnsngiec 120
 QY 121 DSSGTCINPSNMCDGVSHCPGSEDENRCVRLYGNPFILOVYSSQRKSMHPVCODDMNENY 180
 DB 121 dssgtcinpnmcdgvshcpgsedenrcvrllygnpfllqvysqrksmhpvcoddmneny 180
 QY 181 GGAACRDMGYKKNFYSSQGITVDGSGTSPMKLNTSAGNVDIKKLYHSPACSKAVYSIR 240
 DB 181 ggaacrdmgynkfysqgltvdsgstsfmklntsaagnvdlykklyhsadcskavyslr 240
 QY 241 CIACGVNLNLSRQSRIVGESALPGAMPQVSLHVNHVCGSGSITTPMIVTAAHCVEK 300
 DB 241 ciacgvnlinsrqsrtivgesalpgampqvslhvgnhvvcgssiltptvtaahcvek 300
 QY 301 PLNNPWHMTAFAGILRQSFMTYAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDL 360
 DB 301 plnpwhmtafagilrtqsfmtyagqyvekvishpnysdsktkndialmkkpltfndl 360
 QY 361 VPRVCLPNPGMMLQPEQLCWMISGMGATEEKGKTSFVLNAKVLLTETORCNSRYVYDNL 420
 DB 361 vprvcclpnpgmmlqpeqlcwmisgmgateekgtsevlnaakvlltctcnsryydnll 420
 QY 421 TPAMICAGFLQGNVDSGCGDSCGPLVTSKNNITMWLIGDTSMSGCAKAYRPGVYGNVMVF 480
 DB 421 tpamicagflqgnvdsccgdsqgplvtsknnlwmlligdtswsgcakayrpgvygnvmvf 480
 QY 481 TDMYTRQMRADG 492
 DB 481 tdmytrqmradg 492

RESULT 2

AAAY44406
 ID AA444406 standard; Protein: 492 AA.

XX AA444406;

DT 22-MAR-2000 (first entry)

DE Human 20P1F12-GTCC protein.

XX 20P1F12; TMRSS2; androgen; serine protease; 20P1F12-GTCC1; cancer;

KW transmembrane protein; colon; prostate; prostate tumour,

XX Homo sapiens.

XX WO9962942-A2.

PD 09-DEC-1999.

XX

PF 01-JUN-1999; 99WO-US12253.

XX 01-JUN-1998; 98US-0087598.

PR 29-JUN-1998; 98US-0091474.

PR 14-APR-1999; 99US-0129521.

XX (UROC-) UROGENESIS INC.

PA (AFAR/) AFAR D E.

PA (HUBER/) HUBERT R S.

PA (LEONG/) LEONG K.

PA (RAITANO/) RAITANO A B.

PA (SAFE/) SAFEFAN D C.

PI Afar DE, Hubert RS, Leong K, Raitano AB, Safiran DC;

DR WPI: 2000-116363/10.

DR N-PSDB: AAZ29636.

XX Novel cell surface antigen useful to treat colon and prostate cancer -

PS Claim 1; Fig 1; 58pp; English.

The present sequence is the 20P1F12 protein (also known as the TMRSS2 protein) which is a prostate-specific, androgen-regulated, cell surface serine protease. It is a glycosylated type II transmembrane protein with an extracellular C-terminal serine protease domain, a scavenger receptor cysteine-rich domain, an LDL receptor class A domain and a predicted transmembrane domain. Host cells can be transformed to produce this protein, using vector containing 20P1F12/TMRSS2 gene (also designated 20P1F12-GTCC1, as deposited with ATCC accession number 207097).
 CC Anti-20P1F12/TMRSS2 antibodies may be used as therapeutic agent for prostate and colon cancers, to image prostate cancer cells and CC prostate tumours, to identify ligands and cellular constituents that bind to a 20P1F12/TMRSS2 gene product and for use as cancer vaccines.

XX Sequence 492 AA:

Query Match 100.0%; Score 2717; DB 21; Length 492;
 Best Local Similarity 100.0%; Pred. No. 2.9e-203;
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNSGSPPAIGPYENHGYRPNAPQTYVPTVYEVHNAQYRSPVQYAPRVLTQA 60
 DB 1 malnsghppaigpyenhyqpenypaqrvtyevhpqyypvqyaprvltqga 60
 QY 61 SNPVVCTOPKSPSGVCTSKTKKALCITLTLTGTFVGAALAGLLMKFMGSCNSNGIEC 120
 DB 61 snpvcctopkspsgvctsktkkalcitltltgtflvgaalaagllmkfmgscnsngiec 120
 QY 121 DSSGTCINPSNMCDGVSHCPGSEDENRCVRLYGNPFILOVYSSQRKSMHPVCODDMNENY 180
 DB 121 dssgtcinpnmcdgvshcpgsedenrcvrllygnpfllqvysqrksmhpvcoddmneny 180
 QY 181 GGAACRDMGYKKNFYSSQGITVDGSGTSPMKLNTSAGNVDIKKLYHSPACSKAVYSIR 240
 DB 181 ggaacrdmgynkfysqgltvdsgstsfmklntsaagnvdlykklyhsadcskavyslr 240
 QY 241 CIACGVNLNLSRQSRIVGESALPGAMPQVSLHVNHVCGSGSITTPMIVTAAHCVEK 300
 DB 241 ciacgvnlinsrqsrtivgesalpgampqvslhvgnhvvcgssiltptvtaahcvek 300
 QY 301 PLNNPWHMTAFAGILRQSFMTYAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDL 360
 DB 301 plnpwhmtafagilrtqsfmtyagqyvekvishpnysdsktkndialmkkpltfndl 360
 QY 361 VPRVCLPNPGMMLQPEQLCWMISGMGATEEKGKTSFVLNAKVLLTETORCNSRYVYDNL 420
 DB 361 vprvcclpnpgmmlqpeqlcwmisgmgateekgtsevlnaakvlltctcnsryydnll 420
 QY 421 TPAMICAGFLQGNVDSGCGDSCGPLVTSKNNITMWLIGDTSMSGCAKAYRPGVYGNVMVF 480
 DB 421 tpamicagflqgnvdsccgdsqgplvtsknnlwmlligdtswsgcakayrpgvygnvmvf 480

OY 481 TDMYROMRADG 492
 DB 481 tdwlyrqmrads 492

RESULT 3

AAV92050
 ID AAV92050 standard; Protein; 492 AA.

XX AAV92050;

XX 01-AUG-2000 (first entry)

DE HRPc6/7 polypeptide from androgen-inducible gene clone.

KM Androgen inducible; testosterone; prostate cancer; cytostatic;
 KW TMPRSS2; diagnosis.

OS Homo sapiens.

PN W0200018961-A2.

PD 06-APR-2000.

PF 30-SEP-1999; 99WO-US22535.

PR 30-SEP-1998; 98US-0163759.

PR 30-SEP-1998; 98US-0164159.

PA (MILL-) MILLENNIUM PHARM INC.

PI Macbeth KJ, Shyjan AW;

DR WPI: 2000-293182/25.

DR N-PSDB: AAA08803.

PT Novel methods for identifying compounds for treating prostate cancer
 PT comprising measuring the level of expression or activity of 1 or more
 PT of 11 genes or their products

PS Claim 2; Fig 3; 108pp; English.

CC This protein is encoded by a gene which is androgen (e.g. testosterone)
 CC inducible in androgen-dependent prostate cancer cells (e.g. LNCaP cells)
 CC and constitutively expressed in androgen-independent prostate cancer
 CC cells (e.g. LNCaP cells). Agents which decrease the expression or
 CC activity of these clones may slow or arrest the growth of prostate cancer
 CC cells or may kill them. HRPc6/7 can be obtained from the sequence of
 CC the known gene for TMPRSS2. A compound useful for treating prostate
 CC cancer can be identified in a novel method comprising measuring the
 CC expression level, or activity, of HRPc6/7, 3, 6/7, 8, 9, 10, 13, 14, 15,
 CC 19, or peripheral-type benzodiazepine receptor (PBR) in a cell, in the
 CC presence and absence of a test compound. The sequences may also be used
 CC in diagnosis of prostate cancer and to determine efficacy of treatment
 CC for prostate cancer.

SO Sequence 492 AA:

Query Match 99.7%; Score 2709; DB 21; Length 492;

Best Local Similarity 99.8%; Pred. No. 1.2e-202;

Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNCGSPAIPIPIYENHGOPEBNPPAOPVYPTVEVHPAOTYSPVQVAPRLTQA 60

DB 1 malnsgspaiipiyenhyqpenypaqptvtyevhpqyypvqyaprvltqda 60

OY 61 SNPVVCTOKSPSGVCTCTKTKALCTLTGLGFTLVGALAGLLMKFMGSKCSNSGIRE 120

DB 61 snpvcvtokspsgvctctkktkalcitltitglfvgaaagllmkfmgskcsnsgrle 120

OY 121 DSSGTCINSMNCDGVSHCPGDEENRCVRLYGNPFILOYSSQKSMHPVCODDMNENY 180

DB 121 dssgtcinspncdgvshcpbggedenrcvrllygsnfllyssqkrshpvcoddmneny 180
 OY 181 GRAACRDMGCKNNFYSSQGIYDDSGSTSPFKLNTSAGNNDIYKLYHSDACSSKAYVSLR 240
 DB 181 graacrdmgyknnfyssqgiyddsgstsfmk lntsaqnvdlyklyhsdacsakavyslr 240
 OY 241 CIACGVNLNNSRQSRIVGSEALPGAMPVOSLHVONVHVCAGSITPEWIVAAACVEK 300
 DB 241 ciacgvnlnsrsgsrivvgesalpgampvqslhvgnvhwvcgsiltpewivaaacvek 300
 OY 301 PLNNPWHMTAFAGILROSEMFYAGYQVEKVISHPNYSKTKNNDIALMKLQKPLTFNDL 360
 DB 301 plnpwhmtafagilrsgfmfyagayqvekvishpnysdsktkndialmkqlkpltfndl 360
 OY 361 VKPVCLPNEGMMIQPOLCHISGWCATEEKGTSEVLNAAKULLIETQCNSTRYVDNLI 420
 DB 361 vkpvcclpnpqmmliqpeqlcwisgwaateekgtsevlnaakvllletqcnstryvndli 420
 OY 421 TPAMICAGFLOGNVDSGCGSGPLVTSKNNIMWLIIGDTWSGSCAKAYRPGYGVNMYF 480
 DB 421 tpamicagflqgnvdsccgsgplvtsknnimwliigdtwsqscakayrpgygvnmvf 480
 OY 481 TDMYROMRADG 492
 DB 481 tdwlyrqmrads 492

RESULT 4

AAAB36901
 ID AAB36901 standard; Protein; 492 AA.

XX AAB36901;

XX 26-FEB-2001 (first entry)

DE Human TMPRSS2 protein.

KW Prostate specific androgen regulated protein; ARSDRL; TMPRSS2;
 KW PART-1; neoplastic.

OS Homo sapiens.

PN W0200065067-A2.

PD 02-NOV-2000.

PF 21-APR-2000; 2000WO-US10920.

PR 23-APR-1999; 99US-0130778.

PR 30-AUG-1999; 99US-0151585.

PR 30-DEC-1999; 99US-0174003.

PR 24-JAN-2000; 2000US-0177751.

PA (UNITV) UNITV WASHINGTON.

PI Nelson PS, Hood L, Lin B;

DR WPI: 2000-679676/66.

DR N-PSDB: AAC83325.

PT Polynucleotide encoding prostate specific androgen regulated
 PT polypeptides and inhibitor of the peptides useful for treating or
 PT reducing the progression of prostate neoplastic condition in an
 PT individual -

PS Claim 63; Page 86-88; 121pp; English.

CC The present invention relates to prostate specific androgen regulated
 CC proteins. The invention may be used to determine an expression level
 CC of the prostate-specific proteins ARSDRL, TMPRSS2, or PART-1 in a
 CC fluid sample or prostate cell sample from an individual. It may also
 CC be used for diagnosing and predicting the susceptibility of a

CC prostate neoplastic condition in an individual. Inhibitors of the
CC proteins are useful for treating or preventing the progression of a
CC prostate neoplastic condition.

Sequence 492 AA:

Query Match 99.2%; Score 2696; DB 21; Length 492;
Best Local Similarity 98.8%; Pred. No. 1.2e-201;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNSGSPPAICPYENHGYOPENPYPAQIVPVYVEYHPAQYRPSVPQYAPRVLTA 60
DB 1 malnsghppaipyenhyqpenpypaqivpvtyevhpqyypvpqyaprvltga 60
OY 61 SPPVCTOPKSPSGVTCSTKTKKALCITLTGTFVGAALAGLLMKFKMGSKNSGIC 120
DB 61 sppvctopkspsgvctcstkktkalcitltgtflvgaalaagllwmkfgsknsqlec 120
OY 121 DSSGTCINPSNMCDVSHCPGGEDENRVRLYGPNFILQVSSQRKSMHPVCODDMNENY 180
DB 121 dssgtcinpsnmcdivshcpggedenrcvrltgpnfllqvssqrksmhpvcoddmneny 180
OY 181 GRAACRDMGKYNKNTFYSSGIVDDSGTSPMKLNTSAGNVDIKKILYHSDACSKAVSLR 240
DB 181 graacrdmgyknntfysgqivddsgstsfmklnsaagnvdiykkilyhsdacsksavslr 240
OY 241 CIACGVNLNSRQSRIVGSESLPGAMPQVSLHYQNVHVCGSITTPWMTAAHCVEK 300
DB 241 ciacgvnlnsrqsriyvgesalpgampqvslhyqnvhvcgsiltppwmtaahcvek 300
OY 301 PLNNPWHMTAFAGILRQSFMEYAGYQVEKVISHPNYDSKTKNNDIAMKLOKPLTFNDL 360
DB 301 plnpwhmtafagilrqsfmfyagygqvkvshpnydsktkndialmkkqkpltfndl 360
OY 361 VKPVCILPNPMMLOPQOLCMISGKATEEKGKTSVLNAKVLLETQRNRYVDNL 420
DB 361 vkpvcilpnpmmlpbeqlcmisgwateekgktsvlnaakvllletqrnsryyvdnl 420
OY 421 TPAMICAGFLOGNVDSGCGDGGPLVTSKNNIMWMLIGDTSMSGCAKAYRPGVYGNVWF 480
DB 421 tpamicagflggnvdscgdsqgplvtstnnimwlligdtswsgcakayrpgvygnvmf 480
OY 481 TDMYIRQMRADG 492
DB 481 tdmyirymkang 492

RESULT 5
ID AAY57280 standard; Protein: 492 AA.
XX AAY57280;
AC AAY57280;
XX 06-JUN-2000 (first entry)
DE Ovr115 homolog protein.
XX Ovr115 homolog protein.
KW CSC; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;
endometrial; uterine; Lung; cytotoxic.
XX Homo sapiens.
OS Homo sapiens.
XX MO200012758-A1.
PN MO200012758-A1.
PD 09-MAR-2000.
XX 09-MAR-2000.
PF 01-SEP-1999; 99WO-US19655.
XX 01-SEP-1999; 99WO-US19655.
PR 02-SEP-1998; 98US-0098880.
XX 02-SEP-1998; 98US-0098880.
PA (DIAD-) DIADEXUS LLC.
XX (DIAD-) DIADEXUS LLC.

PI Salceda S, Sun Y, Recipon H, Caferkey R;
XX WPI: 2000-256657/22.
DR N-PSDB: AA290478.

PT Diagnosing, staging, monitoring, imaging and treating cancer especially
PT gynecological cancers e.g. breast, ovarian cancer and lung cancer,
PT involves measuring cancer specific gene levels in cells and body fluids

Disclosure: Page 52-54; 58pp; English.

XX The invention relates to detecting, diagnosing metastasis and staging
CC cancer by measuring levels of cancer specific genes (CSG) in cells,
CC tissues or body fluids. Their remission and progression, decreases and
CC increases in CSG levels, is also monitored, by periodic sample analysis.
CC The methods are useful for detecting cancers, especially gynecologic
CC cancers which include ovarian, breast, endometrial and uterine cancer
CC and lung cancer. Antibodies against the CSGs labeled with paramagnetic
CC ions or a radioactive agent is useful for imaging cancer and when conjugated
CC with a cytotoxic agent are useful for treating cancer. The present
CC sequence represents a Ovr115 homolog protein, that can be used
CC for the detection of the various cancers.

Sequence 492 AA:

Query Match 99.2%; Score 2696; DB 21; Length 492;
Best Local Similarity 98.8%; Pred. No. 1.2e-201;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNSGSPPAICPYENHGYOPENPYPAQIVPVYVEYHPAQYRPSVPQYAPRVLTA 60
DB 1 malnsghppaipyenhyqpenpypaqivpvtyevhpqyypvpqyaprvltga 60
OY 61 SPPVCTOPKSPSGVTCSTKTKKALCITLTGTFVGAALAGLLMKFKMGSKNSGIC 120
DB 61 sppvctopkspsgvctcstkktkalcitltgtflvgaalaagllwmkfgsknsqlec 120
OY 121 DSSGTCINPSNMCDVSHCPGGEDENRVRLYGPNFILQVSSQRKSMHPVCODDMNENY 180
DB 121 dssgtcinpsnmcdivshcpggedenrcvrltgpnfllqvssqrksmhpvcoddmneny 180
OY 181 GRAACRDMGKYNKNTFYSSGIVDDSGTSPMKLNTSAGNVDIKKILYHSDACSKAVSLR 240
DB 181 graacrdmgyknntfysgqivddsgstsfmklnsaagnvdiykkilyhsdacsksavslr 240
OY 241 CIACGVNLNSRQSRIVGSESLPGAMPQVSLHYQNVHVCGSITTPWMTAAHCVEK 300
DB 241 ciacgvnlnsrqsriyvgesalpgampqvslhyqnvhvcgsiltppwmtaahcvek 300
OY 301 PLNNPWHMTAFAGILRQSFMEYAGYQVEKVISHPNYDSKTKNNDIAMKLOKPLTFNDL 360
DB 301 plnpwhmtafagilrqsfmfyagygqvkvshpnydsktkndialmkkqkpltfndl 360
OY 361 VKPVCILPNPMMLOPQOLCMISGKATEEKGKTSVLNAKVLLETQRNRYVDNL 420
DB 361 vkpvcilpnpmmlpbeqlcmisgwateekgktsvlnaakvllletqrnsryyvdnl 420
OY 421 TPAMICAGFLOGNVDSGCGDGGPLVTSKNNIMWMLIGDTSMSGCAKAYRPGVYGNVWF 480
DB 421 tpamicagflggnvdscgdsqgplvtstnnimwlligdtswsgcakayrpgvygnvmf 480
OY 481 TDMYIRQMRADG 492
DB 481 tdmyirymkang 492

RESULT 6
ID AAY81492 standard; Protein: 283 AA.
XX AAY81492;
AC AAY81492;

Db	121	kvisphnydstktnmdialmklyqplftndlvkxvclpnpqmmllqpeqlciwsgwatgee	180
QY	390	KGKTSSEVLNAARVLLIETQRCNSRYVYDNLITTPAMICAGFLQGNVDSQCGSDGSPPLYTSK	449
Db	181	kytcesevlnaakvllietqrcnsryyvdmlltppmicagflqgnvdsqgsqgxlytsk	240
QY	450	NNIMWLIGDTSWGSCKAKAYRPGYGVNVMVTTDMYRQMRADG	492
Db	241	nnlwmlligdtswsgcakayrpgygvnmvmtfdwlyqgmradg	283
RESULT 7			
AAB32246	ID	AAB32246 standard; Protein: 454 AA.	
XX	AC	AAB32246;	
DT	11-JAN-2001	(first entry)	
DE		Tumour associated differentially-expressed gene 12 protein sequence.	
XX			
XX		Transmembrane serine protease: TADG-12; chromosome 17; vaccination;	
KM		tumour associated differentially-expressed gene 12; cytotastic; human;	
XX		malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.	
XX			
OS		Homo sapiens.	
XX		WO200052044-A1.	
PN			
PD		08-SEP-2000.	
XX			
XX		02-MAR-2000; 2000WO-US05612.	
PF			
PR		03-MAR-1999; 99US-0261416.	
XX			
PA		(UYAR-) UNIV ARKANSAS.	
XX			
PI		O'Brien TJ, Underwood LJ;	
PI		WPI: 2000-533263/48.	
DR		N-PSDB; AAA93842.	
PT		DNA fragment encoding tumor associated differentially-expressed gene 12	
PT		protein used for diagnosing and treating malignant hyperplasia and	
PT		cancers including ovarian cancer -	
XX			
PS		Claim 3; Figure 4; 11bpb; English.	
XX			
XX		This invention relates to a novel transmembrane serine protease called	
XX		tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is	
CC		located on chromosome 17. Sequences AA93846-A93845 and AAB32246-B32249	
CC		represent human TADG-12 cDNA and their corresponding protein sequences.	
CC		A splice variant of TADG-12 (TADG-12V) leads to a truncated protein	
CC		product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12	
CC		exhibits cytosolic activity, and can be used in vaccines and in gene	
CC		therapy. TADG-12 nucleotide and protein sequences are used in the:	
CC		diagnosis of malignant hyperplasia and cancers of the ovary, breast,	
CC		lung, colon, prostate and other cancers where TADG-12 is overexpressed.	
CC		TADG-12 is particularly used as tumour marker for early disease	
CC		diagnosis. TADG12 proteins or fragments can be used to vaccinate an	
CC		individual with cancer, suspected of having a cancer or at risk of	
CC		getting cancer. Sequences AA93846-A93853 represent PCR primers used for	
CC		amplifying the TADG-12 cDNA sequence, and in the quantitative analysis of	
CC		TADG-12 mRNA. AAB32250 represents a peptide fragment of TADG-12, used to	
CC		create anti-TADG-12 antibodies. Sequences AAB32251-B32369 represent	
CC		TADG-12 peptides which target HLA, and may be used in a vaccine or for	
CC		immune stimulation.	
XX			
XX			
SO		Sequence 454 AA;	
Query Match	32.7%;	Score 888;	DB 21; Length 454;
Best Local Similarity	46.3%;	Pred. No. 3.8e-61;	

FT Modified-site 144
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 148
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 176
 FT /note= "N-glycosylated"
 FT Modified-site 197
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 200
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 231
 FT /note= "casein kinase II phosphorylation site"
 FT Active-site 243
 FT /note= "characteristic of serine protease"
 FT Modified-site 249
 FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT Modified-site 260
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 303
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 351
 FT /note= "protein kinase C phosphorylation site"
 FT Modified-site 360
 FT /note= "tyrosine kinase phosphorylation site"
 FT Modified-site 365
 FT /note= "protein kinase C phosphorylation site"
 FT Active-site 385
 FT /note= "characteristic of serine protease"
 FT
 FT W09936550-A2.
 FT
 FT 22-JUL-1999.
 FT
 FT 12-JAN-1999; 99WO-US00655.
 FT
 FT 16-JAN-1998; 98US-0008271.
 FT
 FT (INCY-) INCYTE PHARM INC.
 FT
 FT Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;
 FT Tang YF, Yue H;
 FT WPI, 1999-430616/36.
 FT N-PSDB; AAX87154.
 FT
 FT Novel human protease molecules useful in the treatment of
 FT developmental disorders and/or cancers
 FT
 FT Claim 1; Page 74-75; 90pp; English.
 FT
 FT The present sequence represents novel human protease HUPM-6, as
 FT deduced from the consensus sequence (see AAX87154) of overlapping
 FT cDNA clones obtained from various libraries. Northern analysis
 FT shows expression of HUPM-6 in gastrointestinal, and male and
 FT female reproductive cDNA libraries. Approximately 65% of these
 FT libraries are associated with neoplastic disorders, and 22% with
 FT the immune response. The invention provides 12 new human
 FT proteases, i.e. HUPM-1 to -12 (see AAX87149-60), and the
 FT polynucleotides encoding them (see AAX87149-60). Also provided are
 FT vectors, host cells and methods for producing HUPM polypeptides,
 FT as well as agonists and antagonists of HUPM. Methods for treating
 FT or preventing cell proliferative disorders and immune disorders
 FT using HUPM or HUPM antagonists are claimed.
 FT
 FT Sequence 435 AA:

Query Match 24.9%; Score 676.5; DB 20; Length 435;
 Best Local Similarity 39.1%; Pred No. 9.8e-45;
 Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;
 QY 133 CDGVSHCPGDEENRCVRLY--GP-----NFILOYSSQSKSWHPVCDDMNENYNG 181

Db 75 cdgdlcdpgedeehcvksfpegpavavrlskdrstlqyldsatgwnfaacfanflea 134
 QY 182 RAACRDMGY--KNNFYSQ-----GIYDDSGSTFSFKMLNTSAGVNDYIKRIYHSDAS 232
 Db 135 etecrcmgyskptrfraveipgqdlvdvelfensqelrmrnsq-----pcl 182
 QY 233 SKAVSLRGIACGVNUNSSROSRIVGESALPGAMPQOYSLHVQNVGCGSIITPEWIV 292
 Db 183 sgslvslhlaagel---kptrvvggeasvdswpqyslqydkhvcgsllldphwv 239
 QY 293 TAAHCVKEPLNPMHTAPAGILR--OSFMEYAGYOVERVI---SHPNYDSKTRNDIAL 348
 Db 240 taahcfrkh-tdfwfkvragsaklgsf---pslavakillilefmpy---pkndial 291
 QY 349 MKIQKPLTNDLVKPYCLPMPGMLOPEOLCWTSGGATFEK-GKTSVNLNAKVLLET 407
 Db 292 mkfqflltsgvtvrcpldfdeelpatpjlwlgwftkqngkmsdlilasvqids 351
 QY 408 ORCNSRYVDNLITTPAMICAGFLQGNVDSQGDGSGPLVTSKNNIMWLIGDTSWGSCKAK 467
 Db 352 ttcnaddayqgevtckmmeaglpesgvdtcggdsgplmgsdq-whvvgivswygcg 410
 QY 468 AYRPGVYGNVMVETDMYRQMRAD 491
 Db 411 pstpgvytkvsaylwnlynwkae 434
 RESULT 12
 AAY72558
 ID AAY72558 standard; Protein; 435 AA.
 XX
 AC AAY72558;
 XX
 DT 02-MAY-2001 (first entry)
 XX
 DE Human seripancrin protein.
 XX
 KW Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
 KW arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
 KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
 KW inflammatory disorder; stroke; cardiovascular disease; gene therapy;
 KW vaccine; cytostatic; cerebroprotective; vulnerary; osteopathic.
 XX
 OS Homo sapiens.
 XX
 PN W0200104141-A2.
 PD 18-JAN-2001.
 XX
 PF 04-JUL-2000; 2000WO-EP06211.
 XX
 PR 12-JUL-1999; 99EP-0113428.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Suendermann B, Hofmann U, Matzku S, Wilbert O;
 XX WPI, 2001-147177/15.
 DR N-PSDB; AAD02556.
 DR
 PT New extracellular serine protease Seripancrin, useful for treating
 PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
 PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant
 PT wound healing
 PT
 PS Claim 2; Page 39-40; 45pp; English.
 XX
 CC The present invention relates to seripancrin polynucleotides, and
 CC polypeptides encoded by them. Seripancrin are members of serine
 CC protease family. This protein contains a transmembrane domain,
 CC a low density lipoprotein (LDL) domain, protease domain and a
 CC scavenger receptor cysteine-rich (SRCR) domain. The LDL and SRCR

CC domains help to define the specificity of seripancrin's intra and
CC intermolecular interactions. The polynucleotides and polypeptides
CC of the invention are useful for treating and diagnosing diseases
CC such as arthritis, chronic obstructive pulmonary disorder (COPD),
CC cancer, osteoporosis, aberrant wound healing, angiogenesis,
CC inflammatory disorders, diabetes, stroke and cardiovascular diseases.
CC Seripancrin genes are useful in chromosome localization studies,
CC as tools for tissue expression studies and also in gene therapy.
CC The polypeptides of the invention are used for identifying agonists
CC and antagonists useful for treating conditions associated with
CC seripancrin imbalance. These polypeptides are also useful as vaccines.
CC The present sequence is seripancrin protein. The
CC seripancrin gene is located on human chromosome 11q22-q23.

XX Sequence 435 AA:

Query Match 24.9%; Score 676.5; DB 22; Length 435;
Best Local Similarity 39.1%; Pred. No. 9.8e-45;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

OY 133 CDGVSHCPGEGEDENCVRLY--GP-----NFILOYSSQKSKMHPVCCDDMNENYGC 181
DB 75 cdgaldcpigedeencvsvfpegapvavrlskdrgtlqyldaatgmfsactdntleala 134
OY 182 RAACRDMGY--KNNFYSSQ-----GIVDSGSTSFMKLNTSAGNVDIYKRLYHSDACS 232
DB 135 etacrqmgysskpftraveigpddldvveltenagelmrmnsg-----pcl 182
OY 233 SKAVVSLKCIACGVNLNNSRQRIYGESALPGANPMOYSLHVNHNVCSSGITPEWYIV 292
DB 183 sglslvslhclacgkal---kcprrvvgaeasvswpwyglqdkqhvcgsglllphwyl 239
OY 293 TAAHCVEKPLNPMWTAFAGILR--OSFMFYCAGYQVEKVI---SHPNVDSKTKNNDIAL 348
DB 240 taahfirh-tlvfnwkvragdsklgsf----pslavakllilefpmv---pkndial 291
OY 349 MKLQRLPLFNLDLVKPVCLPNPGMLQPEQLCWIISGNGATEEK--GRTSEVLAIAKYLLET 407
DB 292 mklqfplrtsgvtrpdlcplfddeeltprplwlgwgftkgqgkmsdlldqasqvlds 351
OY 408 ORCNSRIYVDNLTIPAMICAGFLGQNVDSGCGSGPLVTSKNNIMWLIGDTSMSGCGAK 467
DB 352 trcnaddaygevelctkmccaglpesgvdtcggdsggplmlygsdq-whvvglvswgycgg 410
OY 468 AYRPGVGVNVAVFTDMIXROMRAD 491
DB 411 pslpvytlctksaylmvlynvwkae 434

RESULT 13

ID AAY99417 standard; Protein: 432 AA.

XX AAY99417;

XX 08-AUG-2000 (first entry)

DE Human PRO1570 (UNQ776) amino acid sequence SEQ ID NO:275.

KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
transmembrane; secretion; immunoadhesion; pharmaceutical; screening.

XX Homo sapiens.

XX OS WO200012708-A2.

XX PD 09-MAR-2000.

XX 01-SEP-1999; 99WO-US20111.

XX 01-SEP-1998; 98US-0098716.

PR 01-SEP-1998; 98US-0098749.

PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0098936.
PR 09-SEP-1998; 98US-0099356.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100915.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102309.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.

Wed Sep 26 13:56:53 2001

Job time: 104 sec

us-09-615-285-2.rag

Page 13

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:05 ; Search time 25.88 Seconds

(without alignments)
1448.141 Million cell updates/sec

Title: US-09-615-285-2

Sequence: 1 MALNSGSPRAIGPYENHGY.....VYGNVWYFTDWTYRQMRADG 492

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR.68:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	660	24.3	1035	1 A43090	enteropeptidase (E
2	647	23.8	1019	1 A56318	enteropeptidase (E
3	634	23.3	1034	1 A53663	enteropeptidase (E
4	577.5	21.3	1113	2 JE0315	low-density lipopr
5	569.5	21.0	638	1 KOHUP	plasma kallikrein
6	565.5	20.8	417	1 S00845	hepsin (EC 3.4.21.
7	556	20.5	638	1 KORPL	plasma kallikrein
8	538	19.8	638	1 KOMSPL	plasma kallikrein
9	533	19.6	625	1 KEFHU1	coagulation factor
10	529.5	19.5	416	1 S33777	hepsin (EC 3.4.21.
11	518.5	19.1	812	1 PLMS	plasmin (EC 3.4.21
12	501	18.4	421	1 S11674	acrosin (EC 3.4.21
13	501	18.4	1524	2 T30337	polyprotein - Atri
14	490.5	18.1	415	1 A34170	acrosin (EC 3.4.21
15	489.5	18.0	436	2 JX0172	acrosin (EC 3.4.21
16	486.5	17.9	761	2 JCS759	brain-specific ser
17	483	17.8	418	2 A37344	acrosin (EC 3.4.21
18	477	17.6	790	1 PLUG	plasmin (EC 3.4.21
19	475.5	17.5	431	2 S47538	acrosin (EC 3.4.21
20	473.5	17.4	421	2 S29599	acrosin (EC 3.4.21
21	472.5	17.4	810	2 B30848	plasmin (EC 3.4.21
22	470.5	17.3	437	2 S18407	acrosin (EC 3.4.21
23	470.5	17.3	23	2 I46260	plasmin (EC 3.4.21
24	461	17.0	343	1 A57014	prostasin (EC 3.4.
25	456.5	16.8	460	2 B61545	plasmin (EC 3.4.21
26	455	16.7	812	1 PLBO	plasmin (EC 3.4.21
27	451	16.6	2616	2 A57096	nudel protein prec
28	450.5	16.6	267	2 S40006	trypsin (EC 3.4.21
29	450	16.6	270	2 S56160	mast cell tryptase

30	449	16.5	275	2 S40005	trypsin (EC 3.4.21
31	449	16.5	655	1 A46688	hepatocyte growth
32	448	16.5	277	2 S35340	trypsin (EC 3.4.21
33	447	16.5	276	2 A38654	mast cell proteina
34	446.5	16.4	420	2 A55283	acrosin (EC 3.4.21
35	445.5	16.4	810	1 PLHU	plasmin (EC 3.4.21
36	444.5	16.4	455	2 A61545	plasmin (EC 3.4.21
37	443.5	16.3	266	2 S54146	trypsin (EC 3.4.21
38	442.5	16.3	458	1 S00657	apoptotelin(a) (EC
39	441.5	16.2	274	2 JC4171	trypsin (EC 3.4.2
40	441	16.2	786	1 A47547	serine proteinase
41	439	16.2	273	2 A47246	trypsin (EC 3.4.2
42	438	16.1	274	2 S35339	trypsin (EC 3.4.21
43	437	16.1	275	2 S40007	trypsin (EC 3.4.21
44	436	16.0	242	2 S49489	trypsin (EC 3.4.21
45	436	16.0	456	1 KXBO	protein C (activat

ALIGNMENTS

RESULT 1

A43090

enteropeptidase (EC 3.4.21.9) precursor - bovine

N.Alternate names: enterokinase

C.Species: Bos primigenius taurus (cattle)

C.Date: 10-Sep-1999 #sequence,revision 10-Sep-1999 #text_change 10-Sep-1999

C.Accession: A43090; A48874; A61436

R.Ritamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.

Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A.Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease co

A.Reference number: A43090; MUID:94329561

A.Accession: A43090

A.Molecule type: mRNA

A.Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ

A.Residues: 1-1035 <Kir>

A.Cross-references: GB:009859; MTD:9746410; PIDN:AA0026.1; PID:9746411

A.Experimental source: small intestine

R.Lavallie, E.R.; Rehentulla, A.; Reagle, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.

J. Biol. Chem. 268, 23511-23517, 1993

A.Title: Cloning and functional expression of a cDNA encoding the catalytic subunit o

A.Reference number: A48874; MUID:94043122

A.Accession: A48874

A.Molecule type: mRNA

A.Residues: 801-1035 <LAV>

A.Cross-references: GB:L19663; MTD:9416131; PIDN:AA16035.1; PID:9416132

A.Note: Parts of this sequence, including the amino end of the mature protein, were c

J. Right, A.; Janska, H.

J. Protein Chem. 10, 475-480, 1991

A.Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.

A.Reference number: A61436; MUID:92189715

A.Accession: A61436

A.Molecule type: protein

A.Residues: 801-807,'Y',809-827 <LIG>

C.Comment: The mechanism of association with the membrane of the intestinal brush bor

embrae attachment using a signal-anchor sequence.

C.Comment: Conversion from membrane-bound to soluble forms may involve further proces

C.Complex: mature enteropeptidase is variously reported to contain two (heavy and lig

l)ide linked

C.Function:

A.Description: cleaves propeptide from trypsinogen to produce active trypsin

A.Pathway: intestinal digestive hydrolase cascade

C.Superfamily: enteropeptidase: C1r/C1s repeat homology; LDL receptor ligand-binding

C.Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane prot

F:22-38/Domain: transmembrane #status predicted <TMM>

F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>

F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>

F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:358-520/Domain: MAM homology <MAM>

F:542-647/Domain: C1r/C1s repeat homology <C1R>

F:659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <

F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>

F:801-1030/Domain: trypsin homology <TRY>
 F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin
 F:788-912,926-942,957-972,983-1011/Disulfide bonds: #status predicted
 F:841,892/Active site: His, Asp, Ser #status predicted

Query Match 24.3%, Score 660; DB 1; Length 1035;
 Best Local Similarity 33.9%, Pred. No. 5.7e-40;
 Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

65 VCTQKSPSGVCTSKTKALCITLT-----CTPLVGAALAGLMLKFMKSCNSNG 117
 Db 611 VYTGCPVNDVFSTINRMTVLEFITDNLAKGCFKAFNTTGGVLC-----IPECKEDN 663
 QY 118 IECDSGTCINPSMNCDEGVSHCPGEGEDENRCVRLYG-----PNEILOVYSQKSNH 169
 Db 664 FQC-KDGCIPVLNLCDFEPHCKDGDSEAHCVRLFNGTTDSSGLVQFRIQ-----SIWH 716
 QY 170 PYCQDDNMENGRACRDMGKKNNTSSGGLVDSGSTSPKMLNTSAGNVDIYKKLYHSD 229
 Db 717 VACAEKMTQISDDVLCOLLGLGTG--NSSVPTEFTGGGPPVNLNAPNGSL--LTFESQ 771
 QY 220 ACSSKAVVSLRC--IACGVNLNSSRQ--RIVGGSALPGAMPQVSLHVNHYVCGGSII 286
 Db 772 QCLEDSLILQCNKSCGKKLVTOEVSFKIVGSDSRGACMPVVALYFDQVCGASTLV 831
 QY 287 TPFWITVAHCVEKPLNPMWHTAFAGILROSFMFYGAGVQVE-----KVISHPNYDSKT 341
 Db 832 SRDMLVSAHCHYGNRMESKMKAVILGLHMASNL--TSQIETELIDQIVINPNYKRR 888
 QY 342 KNNIDALMLKLPFLFNLDLVKPVCLPFGMMLQPEOLCIVISGWTGTEKGTSEVLNAK 401
 Db 889 KNNIDAMMLLEKKNVYTYIPICLPEENQVFPFRGISIAQMGALYYQGSTADVLQEDD 948
 QY 402 VLLIETQRCNSRYVNDLITPRAMICAFLOGVNDSCGDSGGPGLVTSKNNIMWLIGDTS 461
 Db 949 VPLSNENCCQOQMPEYN-ITEMNVCAGYEGAGVDSGCGSPGLMCGENNRMLLAGVTSF 1007
 QY 462 GSGCAKAYRPGVYGNVFTDWT 484
 Db 1008 GYOCALPNRPGYARVPRTETWI 1030

RESULT 2
 A56318
 enteropeptidase (EC 3.4.21.9) precursor - human
 N:Alternate names: enterokinase
 C:Species: Homo sapiens (man)
 C>Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
 C:Accession: A56318; B43090
 R:Kilameto, Y.; Veille, R.A.; Donis-Keller, H.; Sadler, J.E.
 Biochemistry 34, 4562-4568, 1995
 A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolyt
 A:Reference number: A56318; MUID:95234679
 A:Accession: A56318
 A:Molecule type: mRNA
 A:Residues: 1-1019 <KIT>
 A:Cross-references: GB:009860; NID:9746412; PIDN:AAC50138.1; PID:9746413
 R:Kilameto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
 A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
 A:Reference number: A43090; MUID:94329561
 A:Accession: B43090
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 749-1019 <KIT>
 A:Cross-references: GB:009860
 C:Comment: The mechanism of association with the membrane of the intestinal brush border
 (cited below) or with amino-terminal myristoylation of the heavy chain.
 A:Gene: GDB:PRSS7
 A:Cross-references: GDB:384083; OMIM:226200
 A:Map position: 21q21-21q21

C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and lig
 ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv
 ducts.

A:Function:

A:Description: cleaves activation peptide from trypsinogen to produce active trypsin

A:Pathway: intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding

C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymoge

F:1-784/Product: enteropeptidase heavy chain #status predicted <HCH>

F:12-38/Domain: transmembrane #status predicted <TRM>

F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:342-504/Domain: MAM homology <MAM>

F:526-631/Domain: C1r/C1s repeat homology <C1R>

F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:768-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <

F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>

F:785-1014/Domain: trypsin homology <TRY>

F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin

F:788-912,926-942,957-972,983-1011/Disulfide bonds: #status predicted

F:841,892/Active site: His, Asp, Ser #status predicted

Query Match 23.8%, Score 647; DB 1; Length 1019;
 Best Local Similarity 35.8%, Pred. No. 4.9e-39;
 Matches 138; Conservative 67; Mismatches 134; Indels 26; Gaps 10;

QY 113 CSNSGIECDSSGTCINPSMNCDEGVSHCPGEGEDENRCVRLYG-----PNEILOVYSQ 164
 Db 643 CKADHFQC-KNECEVPLNLDGHLHCEGSDPADCVDFPFGTTNNGLVFRIQ----- 696
 QY 165 RKSHPVCODDNMENYGRACRDMGKKNNTSSGGLVDSGSTSPKMLNTSAGNVDIYK 224
 Db 697 -SIWTFACAEKMTQISDDVLCOLLGLGTG--NSSKPIFTSDG--PVRKINTAP--DGLHI 750
 QY 225 LYHSDACSSKAVVSLRC--IACGVNLNSSR--QSRVGSALPGAMPQVSLHVNHYVCG 281
 Db 751 LTPSOQCCLQDSILRLQCHHKSGCKRLAODITPKIVGSSNAKEGMPVVALYIGGRLLC 810
 QY 282 GGSITTPFWITVAHCVEKPLNPMWHTAFAGILROSFMFYGAGVQVE-----KVISHPNYDS 339
 Db 811 GASLVSSDMLVSAHCHYGNRMESKMKAVILGLHMASNL--TSQIETELIDQIVINPNYKRR 870
 QY 340 KTKNNIDALMLKLPFLFNLDLVKPVCLPFGMMLQPEOLCIVISGWTGTEKGTSEVLNA 399
 Db 871 RKNDIDAMMLLEKKNVYTYIPICLPEENQVFPFRGISIAQMGALYYQGSTADVLQEDD 930
 QY 400 AKVLLIETQRCNSRYVNDLITPRAMICAFLOGVNDSCGDSGGPGLVTSKNNIMWLIGD 459
 Db 931 ADVPLSNENCCQOQMPEYN-ITEMNVCAGYEGAGVDSGCGSPGLMCGENNRMLLAGV 989
 QY 460 SWGSCAKAYRPGVYGNVFTDWT 484
 Db 990 SFGYKCALPNRPGYARVPRTETWI 1014

RESULT 3
 A53663
 enteropeptidase (EC 3.4.21.9) precursor - pig
 N:Alternate names: enterokinase
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
 C:Accession: A53663
 R:Matsushima, M.; Ichinose, M.; Yahagi, N.; Kakel, N.; Tsukada, S.; Miki, K.; Kurokaw
 J. Biol. Chem. 269, 19976-19982, 1994
 A:Title: Structural characterization of porcine enteropeptidase.
 A:Reference number: A53663; MUID:94327548
 A:Accession: A53663
 A:Molecule type: mRNA
 A:Residues: 1-1034 <MAT>
 A:Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123
 A:Note: parts of this sequence, including the amino ends of three chains isolated fro
 (cited below) or with amino-terminal myristoylation of the heavy chain.

Query Match	23.3%;	Score 634;	DB 1;	Length 1034;
Best Local Similarity	33.3%;	Pred. No. 4.4e-38;		
Matches 147;	Conservative 72;	Mismatches 179;	Indels 44;	Gaps 13

RESULT 4
JE0315
low-density lipoprotein receptor-related protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 18-Aug-2000
C:Accession: JE0315
C:Tomalia, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.

A;ResIdues: 1-111,3 <TOM>

Query Match	21.3%;	Score 577.5;	DB 2;	Length 1113;
Best Local Similarity	33.3%;	Pred. No. 5.9e-34;		
Matches 134;	Conservative 67;	Mismatches 155;	Indels 47;	Gaps 14

RESULT

C;Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999

R; Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.

A;Reference number: A00921; MUID:86243359

A;Accession: A00921

```

:::residues: 1-638 <CH

```

A; Cross-references: GB:M13143; NID:g190262;

Biochemistry 30, 2050-2056, 1991

A;Title: Location of the disulfide bonds

A: Accession: A37939

A; molecule type: protein

260-283, 'X', 285-287-291, 'X', 293-295:314-317, 'X', 319-320:321-324, 'X', 329-333:334-339,

525; 538-551; 562, 'X', 564-567; 573, 'X', 575-576; 578-583, 'X', 585; 592-604 <MCM>

C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into active trypsin, synthesized in the liver, circulates as a noncovalent complex with alpha₁-antitrypsin, and is activated by factor XIIa.

are linked by one or more disulfide bonds.

c/coumestrol: the enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal

C:Accession: A38514; S48202; S48203
R:Degeen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.
Genomics 8, 49-61, 1990
A:Title: Characterization of the cDNA coding for mouse plasminogen and localization of t
A:Reference number: A38514; MUID:91184812
A:Accession: A38514
A:Molecule type: mRNA
A:Residues: 1-812 <DESeq>
A:Cross-references: GB:U04766; NID:g200402; PIDN:AAA50168.1; PID:g200403
R:LiJnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.
Eur. J. Biochem. 224, 863-871, 1994
A:Title: Characterization of the murine plasma fibrinolytic system.
A:Reference number: S48202; MUID:95010076
A:Accession: S48202
A:Molecule type: protein
A:Residues: 20-25 <LIJ>
A:Accession: S48203
A:Molecule type: protein
A:Residues: 22-27 <LIJ>
C:Comment: Plasminogen is synthesized by the kidney and is present in plasma and many other
C:Comment: plasminogen is converted into plasmin by plasminogen activators, both plasminogen
C:Comment: immediately after dissociation from the clot. In the presence of the inhibitor, the activa
C:Comment: inhibitor, the activation involves also removal of the activation peptide.
C:Comment: Stromelysin 1 (see PIR:KCMS1) acts on plasminogen to produce angiotatin. Th
etul in treating solid tumors.
C:Function:
A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va
as the walls of the granafan follicle; also activates the urokinase-type plasminogen act
A:Pathway: fibrinolysis
C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology
C:Keywords: angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein; hydro
F:1-96/Domain: plasminogen-related protein precursor homology <PLPH>
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-812/Product: plasminogen #status predicted <PRO>
F:20-96/Domain: activation peptide #status predicted <APT>
F:20-466/Product: angiotatin #status predicted <AST>
F:97-581,582-812/Product: plasmin #status predicted <AMT>
F:97-581,582-812/Product: plasmin #status predicted <ACH>
F:97-581/Domain: chain A #status predicted <ACH>
F:103-181/Domain: kringle homology <KR1>
F:185-262/Domain: kringle homology <KR2>
F:275-352/Domain: kringle homology <KR3>
F:337-454/Domain: kringle homology <KR4>
F:481-560/Domain: kringle homology <KR5>
F:582-812/Domain: chain B #status predicted <BCH>
F:582-805/Domain: trypsin homology <TRY>
F:49-73,53-61,103-181,124-154,152-176,185-262,188-316,206-245,234-257,275-352,296-335,32
bonds: #status predicted
F:78-79/Cleavage site: Glu-Asn (stromelysin 1) #status predicted
F:136,308/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:465-467/Cleavage site: Thr-Val (stromelysin 1) #status predicted
F:581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental
F:624,667,762/Active site: His, Asp, Ser #status predicted

Query Match	19.1%;	Score 518.5;	DB 1;	Length 812;
Best Local Similarity	32.7%;	Pred. No. 7.6e-30;		
Matches 136;	Conservative 54;	Mismatches 143;	Indels 83;	Gaps 16;
Oy 106 WKFMGSK-CNSNGICEDSSGTCI-----NPSMNCDEVSINC--PGEDENRCVRLYEPN 155				
Db 446 WEYCNLKKCSSTG-----GSVELLPVSOBSPGSDSETCMGNGDKGKTAVTAAG 499				
Oy 156 FILOYVSSQKRSWHPVC-----ODMNENTGRAACRPMG-----YKNFVSSGGIVDDSG 205				
Db 500 TPGCGMAQDEPHRHSIFTPQNPRADELEKNCRNPDGCVNGPCWCTTN-----547				
Oy 206 STSPFKLNTSAGNDYIKLYHSDACSSKAVYSILCIAAGVNLSSRO--SRIVGESAL 263				
Db 548 -----PKLY--DYCDIPICASASSPFCGRKPYEPPKKCPGRVYGGCVAN 589				
Oy 264 PGAWPQVSLHVQ--NVAVCGGSITPEWIVYTAACHCEKPLANNPHWTAFAGLRQSEMF 321				
Db 590 PHSWPMQSLRRTFGHGGTGLTAPMEVLTAAHCKLES--SRPEFYVILGADE-----644				

```

Oy      322 YGAGGVQV-----KVISHPNVDSTKNNIDMLMKLOKLEFNDLVKPCPLNPQMLOPE 376
         |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      645 YIRGLDVOEIVSAKLIEPN-----NRDILALKSRPATITTDYVIACLPSPMYWADR 698
                                         |||||
Oy      377 QLCWSMGATE---EKGRKTEVLNAAKVLLIETORCNSRYYVDNLITTPAMICAGFLGN 433
         :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      699 TICVTIGSGEIGTFEGAGRLKE---ADPLVIEKKVCNRYEYLNNRVKSFTELCAQLAOG 754
                                         |||||
Oy      434 VDSCGSDGGCLPVYSKNINIMLIGDTSGSCAAKAIRPGVYGNAVMTDATIYROMR 489
         |||||
Db      755 VDSCGSDGGCLPVCEKDKYIIQGTSMGLGCARPKNKGYYVRVSREVDMIEREMR 810
                                         |||||

RESULT  12
SI1674
acrosin (EC 3.4.21.10) precursor - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: SI1674; S23499; SI2063; A61022; S03330
R:Keytime: S.; Adham, I.M.; Engel, W
A>Title: Nucleotide sequence and exon-intron organization of the human proacrosin gen
A:Reference number: SI1674; MUID:90306003
A:Accession: SI1674
A:Molecule type: DNA
A:Residues: 1-421 <KEI2>
A:Cross-references: EMBL:X54017; NID:g35582; PIDN:CAA37964.1; PID:g1216165
A>Note: The authors translated the codon AGG for residue 64 as Thr and CTG for residu
R:Vaquez-Levin, M.H.; Reventos, J.; Gordon, J.W.
Eur. J. Biochem. 207, 23-26, 1992
A>Title: Molecular cloning, sequencing and restriction mapping of the genomic sequenc
A:Reference number: S23499; MUID:92331659
A:Accession: S23499
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-421 <VA2>
A:Cross-references: EMBL:M77378
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
R:Keytime: S
submitted to the EMBL Data Library, December 1989
A:Reference number: SI2063
A:Accession: SI2063
A:Molecule type: DNA
A:Residues: 1-225; 'R', 227-421 <KEI2>
A:Cross-references: EMBL:X54017
R:Adham, I.M.; Klemm, U.; Walter, W.M.; Engel, W.
Hum. Genet. 84, 125-128, 1990
A>Title: Molecular cloning of human proproacrosin cDNA.
A:Reference number: A61022; MUID:90128988
A:Accession: A61022
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-63; 'T', 65-225; 'V', 227-267; 'R', 269-421 <ADH>
R:Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
FEBS Lett. 244, 296-300, 1989
A>Title: Primary structure of human proacrosin deduced from its cDNA sequence.
A:Reference number: S03330; MUID:89153568
A:Accession: S03330
A:Molecule type: mRNA
A:Residues: 1-63; 'T', 65-119; 'V', 121-165; 'L', 167-267; 'R', 269-344; 'R', 346-421 <BAB>
A:Cross-references: EMBL:Y00970; NID:g28325; PIDN:CAA68784.1; PID:g28326
C:Genetics:
A:Gene: GDB:ACR
A:Cross-references: GDB:119645; OMIM:102480
A:Map position: 22q13-22pter
A:Introns: 26/2; 94/2; 189/2; 237/3
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine protease; sperm
F.1-19/Domains: signal sequence #status predicted <SIG>
F.20-42/Product: acrosin #status predicted <MAT>
F.20-42/Product: acrosin light chain #status predicted <LCH>
F.43-42/Product: acrosin heavy chain #status predicted <HCH>

```

F:43-285/Domain: trypsin homology <TRY>
 F:302-379/Region: proline-rich
 F:22-210/Binding site: carbohydrate (Aen) (covalent) #status predicted
 F:25-154/Disulfide bonds: #status predicted
 F:29-162/Disulfide bonds: #status predicted
 F:73-89/Disulfide bonds: #status predicted
 F:88-142/240/Active site: His, Asp, Ser #status predicted
 F:177-246/Disulfide bonds: #status predicted
 F:209-225/Disulfide bonds: #status predicted
 F:236-266/Disulfide bonds: #status predicted

Query Match 18.4%; Score 501; DB 1; Length 421;
 Best Local Similarity 40.8%; Pred. No. 6.5e-29;
 Matches 107; Conservative 42; Mismatches 87; Indels 26; Gaps 10;

OY 244 CGVNLSNNSQS--RIVGSGSALPGAMPQVSLHV-----QNVHVGSGSITTEPMITVAH 296
 DB 29 CGLRRPNQGGVRLVGGAAHGMAMVMSLDIFRYSNHRHTCGSLNSRWLTAH 88
 OY 237 C-VERLNNPMHMTAFACILROSFMYG-----AGYQ---VEKYISHPNYDSKTKNNDI 346
 DB 89 CTVGR--NNVHDMRLVFGAK-----ITYGNKKPVKAPLDERYVEKIIHKKYNSATEGNDI 143
 OY 347 ALMKLQKPLTFDLYKPVCLPN-PCGMILQPEOLCMTSGKATEEGK-KTSEVLNAKVL 404
 DB 144 ALVETTPPSISGCRFTIGPGLPHFKAGLPRGSGSCWAGVYEKAPRPSSTIMEARVDL 203
 OY 405 IETGRCNSRYVDNLTITPMICAGFLGQVNDSCGDSGGLV--TSKNIMWLIDDTSMG 462
 DB 204 IDLDLCNSTQMTNGKVPQPLNNVACAGIPVKGIDTCGSGGLMCKDSKESAVYVGTISWG 263
 OY 463 SGCAKAYRPGVYGNVVFETDWI 484
 DB 264 VGCALAKRGRGITYATWPLYLMI 285

RESULT 13
 T30337
 polyprotein - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T30337
 R:Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
 submitted to the EMBL Data Library, March 1998
 A:Description: cDNA cloning of oocyte-specific, a chymotrypsin-like protease released from X
 A:Reference number: Z20829
 A:Accession: T30337
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1524 <YAN>
 A:Cross-references: EMBL:U81290; NID:g2981640; PID:g2981641; PIDN:AAC24717.1

Query Match 18.4%; Score 501; DB 2; Length 1524;
 Best Local Similarity 33.9%; Pred. No. 3e-28;
 Matches 124; Conservative 59; Mismatches 129; Indels 54; Gaps 13;

OY 134 DGVSICPGEDNRCVRLGPNFILOYVSSQKSMHPVCOODMNEVGRACRDGKNN 193
 DB 466 DAVEYDDAEQOLARLCGTLPLPT-SSPENTMLIRFKTMENSY-----PGFKVK 537
 OY 194 F-----YSSQGITVDDSGSTSPMKLNTSAGNDVYKLYHSDACSSKAVVSLRCIACG 245
 DB 538 FSFVPEKKEQSLP--VDDPTISM--LHPRAIALDV-----CG 571
 OY 246 VLNNSR--QSRIVGSEALPGAMPQVSLHVQNVHVGSGSITTEPMITVAHCEVRLN 303
 DB 572 MAPMTPKMWLPRIYVGEASPSMPOVOIFELRTFHEGALISPOWILTAHCHRAA-- 629
 OY 304 NPMHTAAGILROSFMYGAGYVEKYIS--HPNYSKTKNNIDIALMKLOKPLTFDNL 360
 DB 630 EPSYTWIAG--DHNRLNVESTQIRIKITIRIHNDVNSSETYDNDIALLYLEPDLNDF 687

OY 361 VKPVLNPNQMLQPEOLCMTSGKATEEGKKTSEVLNAKVLITETGRCSRYVDNLI 420
 DB 668 VAPVCLPEEEVLPASVCVYTGKGNFTAEEDQPALGLOQLPLIDSLICNTSY-YSEL 746
 OY 421 TPAMICAGFLQG-NVDSQCGDSGGLV-TSKNNIMWLIGDTSMGSKAKAVRPGVYGNV 478
 DB 747 TDHMLCAGFPSSKERDQCGDSGGLVCONKEQPSIVGLVSWEGCGRVSPGYTKVR 806
 OY 479 VFTDWI 484
 DB 807 LEFTWI 812

RESULT 14
 A34170
 acrosin (EC 3.4.21.10) precursor - pig
 N:Alternate names: 53k fucose-binding protein
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A34170; S08994; S02428; S04940; S16657; S02780; S10695; S12968
 R:Baba, T.; Kaishibabara, S.; Matanabe, K.; Itoh, H.; Michikawa, Y.; Kinura, K.; Taked
 J. Biol. Chem. 264, 11920-11927, 1989
 A:Title: Activation and maturation mechanisms of boar acrosin zymogen based on the de
 A:Reference number: A34170; MUID:89308595
 A:Accession: A34170
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-415 <BAH>
 A:Cross-references: GB:S04950; NID:g164702; PIDN:AAA31131.1; PID:g164703
 R:Chehova, D.; Toepfer-Petersen, E.; Zucker, A.; Jonakova, V.
 Biol. Chem. Hoppe-Seyler 371, 317-323, 1990
 A:Title: Is spermminogen a modified proacrosin? Isolation, purification, and partial c
 A:Reference number: S08994; MUID:90253655
 A:Accession: S08994
 A:Molecule type: Protein
 A:Residues: 'X', '18', 'X', '20-25', 'X', '27-32', 'X', '34-38', 'X', '40-50' <CEC>
 R:Toepfer-Petersen, E.; Henschen, A.
 FEBS Lett. 226, 38-42, 1987
 A:Title: Acrosin shows zona and fucose binding, novel properties for a serine protein
 A:Reference number: S02428; MUID:88083633
 A:Accession: S02428
 A:Molecule type: Protein
 A:Residues: 17-32;40-55 <TOB>
 R:Adham, I.M.; Klemm, U.; Maier, W.M.; Hoyer-Fender, S.; Tsaousidou, S.; Engel, W.
 Eur. J. Biochem. 182, 563-568, 1989
 A:Title: Molecular cloning of preproacrosin and analysis of its expression pattern in
 A:Reference number: S04940; MUID:89325301
 A:Accession: S04940
 A:Molecule type: mRNA
 A:Residues: 1-7,9-210,'Q',212-216,'VT',219-346,'A',348-388,390-393,'GN',396,'LVE',399
 A:Cross-references: EMBL:X18844
 A:Note: the authors translated the codon CCT for residue 240 as Ala, GCC for residue
 R:Adham, I.M.
 submitted to the EMBL Data Library, March 1989
 A:Reference number: S16657
 A:Accession: S16657
 A:Molecule type: mRNA
 A:Residues: 1-7,9-210,'Q',212-216,'VT',219-346,'A',348-388,390-398,'KELL' <AD2>
 A:Cross-references: EMBL:X18844; NID:g1867; PIDN:CAA32948.1; PID:g1868
 A:Note: the difference at the carboxyl end is due to a frameshift error
 R:Baba, T.; Michikawa, Y.; Kawakura, K.; Aral, Y.
 FEBS Lett. 244, 132-136, 1989
 A:Title: Activation of boar proacrosin is effected by processing at both N- and C-ter
 A:Reference number: S02780; MUID:89171246
 A:Accession: S02780
 A:Molecule type: Protein
 A:Residues: 17-69 <BA2>
 R:Toepfer-Petersen, E.; Steinberger, M.; von Eschenbach, C.E.; Zucker, A.
 FEBS Lett. 265, 51-54, 1990
 A:Title: Zona pellucida-binding of boar sperm acrosin is associated with the N-termin
 A:Reference number: S10695; MUID:90306316
 A:Accession: S10695

```

FT DOMAIN 151 243 SRCR.
FT DOMAIN 256 492 CATALYTIC.
FT ACT_SITE 296 492 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 445 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 244 365 BY SIMILARITY.
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 465 BY SIMILARITY.
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 492 AA; 53847 MW; 7EAAFPDA18609DDA CRC64;

Query Match 99.2%; Score 2696; DB 1; Length 492;
Best Local Similarity 98.8%; Pred. No. 1.3e-193;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNGSPPAIPYYENHGYOPENYPAQPTVYVYEVHAPQYPPYQYAPRLTQA 60
DB 1 MALNGSPPAIPYYENHGYOPENYPAQPTVYVYEVHAPQYPPYQYAPRLTQA 60
OY 61 SNPVVCTOPKSPSGVCTVSKTKKALCTLTTLCTFLVGAALAGLLMKFGSKCSNGIEC 120
DB 61 SNPVVCTOPKSPSGVCTVSKTKKALCTLTTLCTFLVGAALAGLLMKFGSKCSNGIEC 120
OY 121 DSSGTCINPNSMCDGVSHCPGEGEDENRCVRLYGPNTIILOVSSORKSMHPVCODDMNEY 180
DB 121 DSSGTCINPNSMCDGVSHCPGEGEDENRCVRLYGPNTIILOVSSORKSMHPVCODDMNEY 180
OY 181 GRAACRDGMYKNNFYSSQGIYVDSGSTSEPMKINTSAGNVDIKKLYHSDACSSKAVSLR 240
DB 181 GRAACRDGMYKNNFYSSQGIYVDSGSTSEPMKINTSAGNVDIKKLYHSDACSSKAVSLR 240
OY 241 CIACGVNLTNSHQSRIYVGESALPGAMPQVSLHVONVHVCSSITTPWITYAAHCVEK 300
DB 241 CIACGVNLTNSHQSRIYVGESALPGAMPQVSLHVONVHVCSSITTPWITYAAHCVEK 300
OY 301 PLANNMHTAFAGILROSFMFYGAGYOVOKVISHRYDSKTKNNIDIALMKLQKPLTFNDL 360
DB 301 PLANNMHTAFAGILROSFMFYGAGYOVOKVISHRYDSKTKNNIDIALMKLQKPLTFNDL 360
OY 361 VKPVCILPNPQMLQPEQLCWMISGMATEBEKGTSEVLNAKVLLETQRCNSRYVDNLI 420
DB 361 VKPVCILPNPQMLQPEQLCWMISGMATEBEKGTSEVLNAKVLLETQRCNSRYVDNLI 420
OY 421 TPAMICAGFLQGVNVDSCGDSGGLVTSKNNITWMLIGTSMGSCAKAVRPVYGVNWF 480
DB 421 TPAMICAGFLQGVNVDSCGDSGGLVTSKNNITWMLIGTSMGSCAKAVRPVYGVNWF 480
OY 481 TDMYIRORADG 492
DB 481 TDMYIRORADG 492
OY 481 TDMYIRORADG 492
DB 481 TDMYIRORADG 492

RESULT 2
TMS2_MOUSE STANDARD: PRT: 490 AA.
ID TMS2_MOUSE
AC 09U108; 09UK4; 09Y82;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-) (EPITHELIALIN) (PLASMIC
TRANSMEMBRANE PROTEIN X).
CN TMRSS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMRSS2 in mouse and
human tissues.";
RL J. Pathol. 191:0-0(2000).
[2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT "A novel mosaic serine protease, epitheliasin.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPsin FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; AF199362; AAF97867.1; -
CC EMBL; AF243500; AAF64186.1; -
CC EMBL; AF113596; AAF21308.1; -
CC HSSP; P00761; IAKS.
CC MGD; MGI:1354381; Tmrss2.
CC InterPro; IPR001254; -
CC InterPro; IPR001314; -
CC InterPro; IPR002172; -
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PROSITE; PS00134; TRYPsin_HIS. 1.
CC PROSITE; PS00135; TRYPsin_SER. 1.
CC PROSITE; PS01209; LDLRA_1. 1.
CC PROSITE; PS00068; LDLRA_2. 1.
CC Hydrolase; Serine protease; Transmembrane; Signal-anchor.
KW DOMAIN 1 83
KW TRANSMEM 84 104
FT FT 105 490
FT FT 111 149
FT FT 150 242
FT FT 254 490
FT FT 294 433
FT FT 343 445
FT FT 439 465
FT FT 112 125
FT FT 119 138
FT FT 132 147
FT FT 132 147
FT FT 243 365
FT FT 279 295
FT FT 408 424
FT FT 435 463
FT FT 111 111
FT FT 212 212
FT FT 474 474
FT FT 122 122
FT FT 178 178
FT FT 320 320
FT FT 474 474
FT FT 474 474
SO SEQUENCE 490 AA; 53479 MW; 07D2B03E4A8DA1A9 CRC64;

Query Match 77.9%; Score 2117; DB 1; Length 490;

```



```

Query Match Similarity 32.28; Score 875.5; DB1; Length 454;
Best Local Similarity 45.98; Pred. No. 4,2e-58;
Matches 189; Conservative 53; Mismatches 149; Indels 21; Gaps

Oy 89 LTLGLFVALLAAGLAGLMKFMGSKCSNSGIECDSGTCINPNSMCDGVSHCPGDEENRC 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52 IYIGIIIALIILALIGLGIHF---DCSGK-YRCHSSFRCIEELIARCGVSDCKDGEYRC 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 149 VRLGPNRIILQVYSSOKRSHMPCVODDMNENYGAACRDMGKYNNRYSQGLYDDSGRS 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 VRVGQNNVQLQVETA--ASMTKMSDMDWKHYVNVACADLGP- SYVSSDNLRYVSLSEQ 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 209 FMKLTSAGNY---DIYKKLYHS---DACSRAVYSLRCIAGVNLNSRSRQIVGES 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 FREEVYSIDLHLPPDQKXTALHSHYVNEGASGHVYTLQCTACGHHRGVS--SRIVGGM 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 262 ALPGAMQWQVSLHGVNVHCGGSIITPEWIVTAHACEKRPKNPNMHWTAFCILKQSFNF 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 SLLSQMPQVQASLQCGYHLCGGSVITPLMITTAHCV-YDLYLPRKSMYIQVGLV--SLD 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 322 YGA-GYQYEVKISHPNDVSKTKNNDAIMLKQPLFNFDLVKRVCLPYPGMMLQPEQCLM 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 NPAPSHLVEKLYVYSKKPRKRLGNDIALMLKLAGPLTFNEKIQVCLPNSSENFPPQGVCM 339
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 381 ISGNGATEE-KGKTSEYLNAAKVVLLIETORCNSRYYYDNLITPAMICAGFLQGVNDSCG 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 TSGMGATEDEGAGDASPYLNHAAVPLISNKICNIRDYVGGIISPSMLCAGYLTVGVDSCG 399
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 440 DSGGPLYTSKNNIMWLLIGDTSWGSGCKAKARPCVYGNVAFYDMYIKORAD 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 DSGGLVLCQERRLMKLVGATSFGICAEVKNPCVYTVTSFIDMTHQEMERD 451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT
4
ENTK_BOVIN
ID PK6072; STANDARD; PRT; 1035 AA.
AC P6072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 33, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRS7 OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Duodenum;
RA MEDLINE=94329561; PubMed=8032624;
RA Kilamco Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinas, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN [2]
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=94043122; PubMed=8226855;
RA Lavallie E.R., Rehemtulla A., Rele L.A., Diblasio E.A.,
RA Ferenc C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinasin.";
RL J. Biol. Chem. 268:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RC TISSUE=Intestine;
RA MEDLINE=92189715; PubMed=1799406;
RA Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinasin.";
RL J. Protein Chem. 10:475-480(1991).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PREENZYMES (TRYPsin, CHYMOTRYPSIN AND CARBOXYPEPTIDASE

```

CC	A) IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPsin WHICH IN
CC	TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC	PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC	-1 CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC	TRYPsinOGEN.
CC	-1 SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC	MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC	-1 SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC	-1 TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC	-1 PIM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC	CLEAVED BY A TRYPsin-LIKE PROTEASE.
CC	-1 SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC	-1 SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1 SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC	-1 SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC	-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC	TRYPsin FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: U09859; AAB40026.1; -;
DR	EMBL: L19663; AAL16035.1; -;
DR	PIR: A61436; A61436.
DR	HSSP: P00763; LDPO.
DR	MEROPS: S01.156; -;
DR	InterPro: IPR000082; -;
DR	InterPro: IPR0000859; -;
DR	InterPro: IPR0000998; -;
DR	InterPro: IPR001190; -;
DR	InterPro: IPR001254; -;
DR	InterPro: IPR001314; -;
DR	InterPro: IPR002172; -;
DR	PIfam: PF00643; CUB; 2.
DR	PIfam: PF00629; MAM; 1.
DR	PIfam: PF01390; SEA; 1.
DR	PIfam: PF00530; SRCR; 1.
DR	PIfam: PF00057; LDL_recept_a; 2.
DR	PIfam: PF00089; trypsin; 1.
DR	PRINTS: PR00722; CHYMOTRYPSIN.
DR	PROSITE: PS00134; TRYPsin_HIS; 1.
DR	PROSITE: PS00135; TRYPsin_SER; 1.
DR	PROSITE: PS01180; CUB; 2.
DR	PROSITE: PS00740; MAM.1; 1.
DR	PROSITE: PS0060; MAM.2; 1.
DR	PROSITE: PS01209; LDLRA_1; 2.
DR	PROSITE: PS0068; LDLRA_2; 2.
KW	Signal-anchor; Glycoprotein; Myristate; Hydroxylase;
KW	Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing.
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 650
FT	DOMAIN 657 695
FT	DOMAIN 694 787
FT	SRCR 787
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 650
FT	DOMAIN 657 695
FT	DOMAIN 694 787
FT	SRCR 787
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 650
FT	DOMAIN 657 695
FT	DOMAIN 694 787
FT	SRCR 787
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 650
FT	DOMAIN 657 695
FT	DOMAIN 694 787
FT	SRCR 787
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 650
FT	DOMAIN 657 695
FT	DOMAIN 694 787
FT	SRCR 787
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 650
FT	DOMAIN 657 695
FT	DOMAIN 694 787
FT	SRCR 787
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 650
FT	DOMAIN 657 695
FT	DOMAIN 694 787
FT	SRCR 787
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 650
FT	DOMAIN 657 695
FT	DOMAIN 694 787
FT	SRCR 787
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 650
FT	DOMAIN 657 695
FT	DOMAIN 694 787
FT	SRCR 787
FT	CHAIN 1 800
FT	CHAIN 1 1035
FT	TRANSMEM 19 47
FT	DOMAIN 197 238
FT	DOMAIN 240 350
FT	DOMAIN 358 520
FT	DOMAIN 540 65

```

FT DISULFID 788 912 INTERCHAIN (BY SIMILARITY).
FT DISULFID 826 912 BY SIMILARITY.
FT DISULFID 926 993 BY SIMILARITY.
FT DISULFID 957 972 BY SIMILARITY.
FT DISULFID 963 1011 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 864 864 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 166 192 MISSING (IN SHORT ISOFORM).
FT CONFLICT 808 R -> Y (IN REF. 3).
SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;

```

```

Query Match 24.3%; Score 660; DB 1; Length 1035;
Best Local Similarity 33.9%; Pred. No. 1.2e+41;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

```

```

QY 65 VCTQKPSGSGVCTSKTKALCITLTL-----GTFVGAALAGLMLKMGSKCSNSG 117
DB 611 VYTGPGVADVSTNRMVTLFTIDMLAKOGFKANFTTGTGIGL-----IPECKEDN 663
QY 118 IICDSSGTCINSNMCDGYSHCPCGDEKRCVRLG-----PFIIOYVSSOKKSH 169
DB 664 FQC-KDGEICPLVNLCDGPGHCKDGSDEAHCVLLENGTSSGLVQFRIO-----SIWH 716
QY 170 PYCDDMNENYRAACRDGYKNNFYSSOGIVDSSGTSFMMKLSAGVNDIYKKLYHSD 229
DB 717 VACAEMWTQIDDDVCQLGLGTG--NSSVPTFSTGGGYVNLNTPNSLSL---LTPSQ 771
QY 230 ACSSAAVSLRC--IACGVNLSSROS-RIVGESALPGAMPQVSLHAVQNVHVCGSII 286
DB 772 OCLEDSLILLOCNYSKCGKLVTOEVSPTKVGSDREGAMPVVALYFDDOVCASLIY 831
QY 287 TPEMTYTAHCVKEKPLNNMHTAFAGILROSFMTFGAIOYE-----KVISHPNDSKT 341
DB 832 SMDMLVSAHCVYGRNMESKMAVALGLHMASVL--TSPTLETRLIDIVINPHNKR 888
QY 342 KNNIDALMLKQPLFENDLVKPCPLNPMMLQPEOLQWISGATEEGKTSSEVLNAK 401
DB 889 KNNIDAMHMLEKAVNTVDIOPICLPEENOVPPRIGICIAAGALIIYGSADVLQED 948
QY 402 VLLITQRCNSRYVYDNLITPAMICAGFLQGVNDSQDGGSGPLVTSKNNIMVLIGDTSW 461
DB 949 VLLITSEKQDQOMPEYN-ITENNVACAGYEGAGVDSQDGGSGPLMKOEENRMLLAGVTSE 1007
QY 462 GSGCAKAVRPGYGVNMTFTMI 484
DB 1008 GYOCALPNRPGVARYPRFTEMI 1030

```

```

RESULT 5
ENTR_HUMAN ID ENTR_HUMAN STANDARD; PRT: 1019 AA.
AC P98073;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

```

```

DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Duodenum;
RX MEDLINE=95234679; PubMed=7718557;
RA Kitamoto Y., Velle R.A., Donis-Keller H., Sadler J.E.;
RT "cDNA sequence and chromosomal localization of human enterokinase,
RL the proteolytic activator of trypsinogen.";
RN [2]
RP SEQUENCE FROM N.A.
RA Holzinger A., Buck C., Maier E.M., Mayerhofer P.U., Roscher A.A.,
RA Sadler J.E., Hadorn H.B.;
RT "Genomic organization of the human enteropeptidase.";
RN [3]
RP Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Blecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehnach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 2L.";
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 749-1019 FROM N.A.
RC TISSUE=Duodenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RL protease composed of a distinctive assortment of domains.";
RN [5]
RP Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
CC MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

FT	DISULFID	910	977		BY SIMILARITY.
FT	DISULFID	941	956		BY SIMILARITY.
FT	DISULFID	967	995		BY SIMILARITY.
FT	CARBOHYD	116	116	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	147	147	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	179	179	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	328	328	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	335	335	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	388	388	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	440	440	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	470	470	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	503	503	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	534	534	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	630	630	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	682	682	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	706	706	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	725	725	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	848	848	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	887	887	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	909	909	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CARBOHYD	949	949	N-LINKED (GLCNAC . .)	(POTENTIAL).
FT	CONFLICT	134	134	O -> E (IN REF. 3).	
FT	CONFLICT	732	732	S -> P (IN REF. 3).	
FT	CONFLICT	754	771	SQOQLDFSLRLQCNHKS ->	RMAKNEDALSPILIA
FT	SEQUENCE	1019 AA;	112923 MW;	B6AA4A25f6d4A563 CRC64;	
(IN REF. 3).					
Query Match 23.8%; Score 647; DB 1; Length 1019;					
Best Local Similarity 35.8%; Pred. No. 1.1e-40;					
Matches 138; Conservative 67; Mismatches 134; Indels 26; Gaps 10;					
QY	113 CNSNGIECDSSGTCINPNMCDGVSHCPGEDEMRVLYG-----PNFILOYVSSQ	164			
	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :				
DB	643 CKADEFQC-KINGECYPLVNLCGDHLHCEGSDSEADCAVFNFNGTNNGLVFRPIQ-----	696			
QY	165 RKSHPVCODDMENYIGRAACRDMGCKNKFVSOGIVDSGSTSMKNTSAGVDITKK	224			
	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :				
DB	697-SIWTACAEMTTTOISNDVCCOLLIGSG-NSSKPIFFSTDG-PVKLNTAP--DGHLI	750			
QY	225 LYHSNACSSKAVALSLRC--IACGNVLNSR-QSRIVGESALPGAMPQVSLHYONVHC	281			
	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :				
DB	751 LTPSOCCLODSLIRLOCNHKSCGKKLAADDTPKIVGSNAKEGMWPVVGVYGGRIIC	810			
QY	282 GGSITTPPEVIYAHCVEKPLNNPMHTAFAGILQSFMF-YGAGYOVERVISHPNYS	339			
	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :				
DB	811 GASLYSSDWLYSAACGYGRNLPSRKWIALGLHKMSULTSPQYPRILDELVINPHNR	870			
QY	340 KTKNNDIAMLKLFKLPFLTNLDLVKPYCLNPFGMTLOPEOLCWSIGNGATEEKGKTSEVINA	399			
	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :				
DB	871 RKRDIDIAMMHLEFKVNYTYDITOPICLDEBNOFPFPGHNCSIAAGCVAVVOGTIANILOE	930			
QY	400 AKVLLIEFORCHRSRVYDNLTIPPAICAGFLQGNVDSQGSGSGFLVYSKNNIMWLIGDT	459			
	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :				
DB	931 ADVPLLISNERCOQMPEYN-ITENNICAGEYGEGIDSCGDSGGELMCQENNRFLAGVT	989			
QY	460 SWGSGCAKAYRPVGYNVMFTDWI	484			
	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :				
DB	990 SFGYKCALPNRGYIARYARSRTIEWI	1014			
RESULT 6					
ENTK_MOUSE STANDARD; PRt: 1069 AA.					
AC	P97435;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-OCT-2000 (Rel. 40, Last annotation update)				
DE	ENTEROPEPTIDASE (EC 3.4.21.9) (EMTEOKTIASAE).				
GN	PROST OR ENTK.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				

NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6; TISSUE=Duoenum;
MEDLINE=98147142; PubMed=9486188;
RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
RT Structure of murine enterokinase (enteropeptidase) and expression in
small intestine during development.";
RL Am. J. Physiol. 274.G342-G349(1998).
CC - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
CC - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
TRYPSINOGEN.
CC - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
SIMILARITY).
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
CC - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPsin FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

DR EMBL: U73378; AAB37317.1; -
DR HSSP: P00763; IDPO.
DR MEROPS: S01.156; -
DR MGD: MGI:1197523; Prss7.
DR InterPro: IPR000082; -
DR InterPro: IPR000859; -
DR InterPro: IPR000998; -
DR InterPro: IPR001190; -
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -
DR InterPro: IPR002172; -
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF01390; SEA; 1.
DR Pfam: PF00530; SRCR; 1.
DR Pfam: PF00057; 1dl_recept_a; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS50060; MAM_2; 1.
DR PROSITE: PS50109; LDLRA_1; 2.
DR PROSITE: PS50068; LDLRA_2; 2.
KW Signal-anchor: Glycoprotein; Myristate; Hydrolase;
KW Serine protease; Zymogen; Transmembrane; Repeat.
FT CHAIN 1 829 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT TRANSMEM 830 1069 CATALYTIC CHAIN (LIGHT CHAIN).
FT DOMAIN 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 227 268 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 270 379 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 387 549 CUB.
FT DOMAIN 569 679 MAM.
FT DOMAIN 686 724 CUB.
FT DOMAIN 723 816 LDL-RECEPTOR CLASS A 2.
FT DOMAIN SRCR.

FT	ACT_SITE	874	874	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	925	925	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1021	1021	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID	2	2	MYRISTATE (POTENTIAL).
FT	DISULFID	229	242	BY SIMILARITY.
FT	DISULFID	236	255	BY SIMILARITY.
FT	DISULFID	249	266	BY SIMILARITY.
FT	DISULFID	668	700	BY SIMILARITY.
FT	DISULFID	695	713	BY SIMILARITY.
FT	DISULFID	707	722	BY SIMILARITY.
FT	DISULFID	817	945	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	859	875	BY SIMILARITY.
FT	DISULFID	959	1027	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1017	1045	BY SIMILARITY.
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	197	197	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	212	212	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	373	373	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	380	380	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	433	433	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	515	515	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	579	579	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	675	675	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	727	727	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	751	751	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	770	770	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	791	791	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	897	897	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	936	936	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	999	999	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	1069 AA:	118735 MW:	E62549EA63743C3D CRC64;

Query Match 23.5%; Score 638; DB 1; Length 1069;
Best Local Similarity 30.5%; Pred. No. 5,3e-40;
Matches 164; Conservative 92; Mismatches 196; Indels 86; Gaps 21;

QY	18	HGYQENPYPAPQTVVPT-----YEV-HQAQYPPSPVQYARVLTQAS----	61
DB	544	NGICSPYP-EPTLVTPPELPDGGPFELMEBNSTFS--PMPDPKYPNQAACIWN	600
QY	62	-----NPV-----VCIDPKSPSGNVCSTKTKAL	85
DB	601	LNARCKNIOIHFQEDDENINDVVEYRGGEDSLILAVYIGP-GPVADISTIRKRTV	659
QY	86	CITLIGTFLVG--AALAAGLIMKFMG--SKCSNSGIECDSSGTCINPNSMCDGVSHPG	141
DB	660	IFTTNNETRRKGFKANFTSGY--YGIPEPCDDDFQC-KDNCIPLNLDSDYRHCBD	715
QY	142	GEDENRCVRLY----GNPFLQYYSQKRSWHPVCODDMNENTGRACADMGYKKNFTSS	197
DB	716	GSDASCVRLNGLTNRNNGLYOF--NIHSIMHATCAENMTTOISNEVCHLIGLS--ANS	771
QY	198	QGVLDVSGSTSPFKLNTSAGNVDIYKKLYHSDACSKAVVSLRC--IAGVNLNSROS-	254
DB	772	SMFISSTGGGPFYRVQAPRGLI---LTPSLQCSODSLILQCNHKSCEKKVYTKVSP	828
QY	255	RIVGESALPGAMPQVSLHYQNVH---VCGGSITTPETVITAAHCVKRPPLNPNMHTA	310
DB	829	KIYGSDAQAGAMPVVALYHNRDSTRLLCASVLSVMSLVSAACHVYRRLNDPTRMTA	888
QY	311	FAGILQSFMEFYAGYQ--VEKYISHPNVDSKTKNNDIALMKLQKPLTNDLYKPCLEN	368
DB	889	VGLHMQSNLTPDQVVRVAVDQIVINPHYDRRKRVNDIAMHLEFVNVNTDVIQPCLE	948
QY	369	PGMILQPEOLCWTISGATE-EKGTSEVLNNAKVLLIFETORNSRYVDNLITPMICA	427
DB	949	ENQIFIPGRCSLAGGYDKINAGSTVDYLKADVPLNISNCKQOQDLPXYN-ITSMICA	1007
QY	428	GLQGNVDSQCGDSGGLVTSKNNIMWLIGTWSGSCAKAVRPGVYGVNVMFTDNIY	485
DB	1008	GYEGGIDSCQDGSGLPMCKQENNRWFLVGVTSFVQCALPNHPRGVYVAVSQFIEMH	1065

ENTK_PIG	STANDARD:	PRT:	1034 AA.
DR	ENTK_PIG		
AC	P98074:		
DT	01-FEB-1996 (rel. 33, Created)		
DT	01-FEB-1996 (rel. 33, Last sequence update)		
DT	15-JUL-1998 (rel. 36, Last annotation update)		
DE	ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).		
GN	PRSS7 OR ENTK.		
OS	Sus scrofa (Pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
OX	NCBI_TaxId=9823;		
RA	[1]		
RA	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC	TISSUE=Duodenal mucosa;		
RX	MEDLINE=94327548; PubMed=8051081;		
RA	Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S.,		
RA	Miki K., Kurokawa K., Tashiro K., Shokawa K., Shinomura K.,		
RA	Umayama H., Inoue H., Takahashi T., Takahashi K.;		
RT	"Structural characterization of porcine enteropeptidase.";		
RL	J. Biol. Chem. 269:19976-19982(1994).		
CC	- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC		
CC	PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE		
CC	A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN		
CC	TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,		
CC	PROCARBOXYPEPTIDASES, AND PROELASTASES.		
CC	- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LEU- -ILE-7 BOND IN		
CC	TRYPSINOGEN.		
CC	- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN		
CC	(HEAVY) CHAIN, AND A MINI CHAIN.		
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLY).		
CC	- NOTE: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS		
CC	CLEAVED BY A TRYPSIN-LIKE PROTEASE.		
CC	- PFM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.		
CC	- SIMILARITY: CONTAINS 2 LDI-RECEPTOR CLASS A DOMAINS.		
CC	- SIMILARITY: CONTAINS 2 CUB DOMAINS.		
CC	- SIMILARITY: CONTAINS 1 SRCR DOMAIN.		
CC	- SIMILARITY: CONTAINS 1 MAM DOMAIN.		
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE		
CC	TRYPSIN FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-slb.ch/announce/		
CC	or send an email to license@lsb-slb.ch).		
CC	-----		
DR	EMBL; D30799; BA06459.1; -.		
DR	HSSP; P00763; IDPO.		
DR	MEROPS; S01.156; -.		
DR	InterPro: IPR000082; -.		
DR	InterPro: IPR000859; -.		
DR	InterPro: IPR000998; -.		
DR	InterPro: IPR001190; -.		
DR	InterPro: IPR001254; -.		
DR	InterPro: IPR001314; -.		
DR	InterPro: IPR002172; -.		
DR	Pfam: PF00431; CUB; 2.		
DR	Pfam: PF00629; MAM; 1.		
DR	Pfam: PF01390; SEA; 1.		
DR	Pfam: PF00530; SRCR; 1.		
DR	Pfam: PF00057; ldl_recept_a; 2.		
DR	Pfam: PF00089; trypsin; 1.		
DR	PRINTS; PR00722; CHYMOTRYPSIN.		
DR	PROSITE; PS00134; TRYPSIN_HIS. 1.		
DR	PROSITE; PS00135; TRYPSIN_SER; 1.		
DR	PROSITE; PS01180; CUB; 2.		
DR	PROSITE; PS00740; MAM_1; 1.		

DR	PROSITE: P550060; MAM_2; 1.
DR	PROSITE: PS01209; LIDRA_1; 2.
DR	PROSITE: PS50068; LIDRA_2; 2.
KW	Signal-anchor; Glycoprotein; Myristate; Hydrolase;
KW	Serine protease; Zymogen; Transmembrane; Repeat.
FT	CHAIN 52 117
FT	CHAIN 118 799
FT	CHAIN 800 1034
FT	CHAIN 19 47
FT	TRANSMEM 197 238
FT	DOMAIN 240 349
FT	DOMAIN 357 519
FT	DOMAIN 539 649
FT	DOMAIN 656 694
FT	DOMAIN 693 786
FT	ACT_SITE 840 840
FT	ACT_SITE 881 881
FT	ACT_SITE 966 986
FT	ACT_SITE 2 2
FT	LIPID 199 212
FT	DISULFID 206 225
FT	DISULFID 219 236
FT	DISULFID 658 670
FT	DISULFID 665 663
FT	DISULFID 677 692
FT	DISULFID 787 911
FT	DISULFID 825 841
FT	DISULFID 925 992
FT	DISULFID 966 971
FT	DISULFID 982 1010
FT	CARBOHYD 116 116
FT	CARBOHYD 147 147
FT	CARBOHYD 170 170
FT	CARBOHYD 194 194
FT	CARBOHYD 283 283
FT	CARBOHYD 343 343
FT	CARBOHYD 350 350
FT	CARBOHYD 403 403
FT	CARBOHYD 455 455
FT	CARBOHYD 485 485
FT	CARBOHYD 518 518
FT	CARBOHYD 549 549
FT	CARBOHYD 645 645
FT	CARBOHYD 697 697
FT	CARBOHYD 701 701
FT	CARBOHYD 721 721
FT	CARBOHYD 740 740
FT	CARBOHYD 761 761
FT	CARBOHYD 804 804
FT	CARBOHYD 863 863
FT	CARBOHYD 902 902
FT	CARBOHYD 964 964
FT	CARBOHYD 1034 AA; 114776 MW; 038BC64CF64C368 CRC64;
FT	SEQUENCE
Query Match	23.3%; Score 634; DB 1; Length 1034;
Best Local Similarity	33.3%; Pred No. 1e-39;
Matches 147; Conservative	72; Mismatches 179; Indels 44; Gaps 13.
65	VCTQPKSPSGIVCTSKTKKALCIT--LTIG---TFIVGAALAGLIMKFMGSKCSNSG 117
Db	610 VYTGPGPVEDVSTNIRMTVLFTNDALTKGGEKFAKFTTGYHLG-----IPECKEDN 662
Qy	118 IICDSSGTCINSNMCDGVSHCPGGEDEMKCYRLIG--PNFLILOYSSQRKSMHPVCODD 175
Db	663 FQCE--NGECVLILVNLCDGFSHCKDSDSEHCVPLFNGLTANNSGLVQFRIQSITWHTACAN 721
Qy	176 WNEENTGRAACRDMGKKN-----FYSSQIGIVDSCGTSFMKNTSAGNVDIKKIYLHSDA 230
Db	722 WTTQTSDDVCQLGLGTGCTGSSMPFS-----SGGGPVPKVLNTAPNGSLI---LTASD 771
Qy	231 CSSKAVVSLRC--IACGVNLSSROS-RIVGESALPGAMPQVSLHVNQVHVGCSIT 287

```

Db 772 CREDLILLQCNHSCGCKROVAQEVSPKIVGNDSEHAGMPWVALYNGOLICGASLVS 831
OY 288 PEMIYTAACHVEKPLNPNHMTAFAGILRQSEMFYAGYO-----VEKISHPNVSKRK 342
Db 832 RDMLVSAACHVGRNLEPCKRAKILGLHMTSNL---TSPQIYTRLDELIVINHTRRRK 888
OY 343 NNDIALMLKQKPLTFENDLVKPYCLPNPMMLOPEOLCMTISGMGATEKCKTSEVLAAKV 402
Db 889 DSDIAMMHLEFVNYNDYIQIPCLPEENQVFPGRICSTIAGNCKYVQSGPADIIQEAUV 948
OY 403 LLIETORCSRYVNDLITPAMICAGFLOGNDSCGDSGGLVTSKNNIMWILIGDTSWG 462
Db 949 PLISNKKCCQOMPEYN-IENNMCAAGEEGIDSCGDSGGLMCIENNRMLAGVTSFG 1007
OY 463 SGCAKAYRPGVYGNWVFETDWI 484
Db 1008 YOCALPNRPVYARVAFKTEWI 1029

RESULT 8
KAL_HUMAN STANDARD: PRT: 638 AA.
AC P03952:
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update).
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININOGENIN) (FLETCHER FACTOR).
GN KLK3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86243359; PubMed=3521732;
RX Chung D.W., Fujikawa K., McMullen B.A., Davie E.W.;
RX "Human plasma prekallikrein, a zymogen to a serine protease that
RX contains four tandem repeats."
RL Biochemistry 25:2410-2417(1986).
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=91152016; PubMed=1998666;
RX McMullen B.A., Fujikawa K., Davie E.W.;
RX "Location of the disulfide bonds in human plasma prekallikrein: the
RX presence of four novel apple domains in the amino-terminal portion of
RX the molecule."
RL Biochemistry 30:2050-2056(1991).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIa, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- DISEASE: DEFECTS IN KLK3 ARE THE CAUSE OF FLETCHER FACTOR
CC DEFICIENCY, A BLOOD COAGULATION DEFECT.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
CC EMBL, M13143; AAA60153.1; -
CC DR PIR: A00921; KOHUP.

```

```

DR PIR: A37939; A37939.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.212; -.
DR MIM: 229000; -.
DR InterPro: IPR00177; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR InterPro: IPR003014; -.
DR Pfam: PR00024; PAN. 4.
DR Pfam: PR00089; trypsin. 1.
DR PRINTS: PR00005; APPLEDOMAIN.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_SER. 1.
DR PROSITE: PS00135; TRYPSIN_SER. 1.
DR PROSITE: PS00495; APPLE. 4.
KW Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 CATALYTIC.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
FT ACT_SITE 77 77 CHARGE RELAY SYSTEM.
FT DISULFID 21 104
FT DISULFID 47 77
FT DISULFID 51 57
FT DISULFID 111 194
FT DISULFID 137 166
FT DISULFID 141 147
FT DISULFID 201 284
FT DISULFID 227 256
FT DISULFID 231 237
FT DISULFID 292 375
FT DISULFID 318 347
FT DISULFID 322 328
FT DISULFID 340 345
FT DISULFID 383 503
FT DISULFID 419 435
FT DISULFID 517 584
FT DISULFID 548 563
FT DISULFID 574 602
SQ SEQUENCE 638 AA; 71369 MW; E62P9C1053838FB4 CRC64;

Query Match 21.0%; Score 569.5; DB 1; Length 638;
Best Local Similarity 32.8%; Pred. No. 3.7e-35;
Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;

OY 55 RVLTOASNPVYCTQPKSPSGTQVCTSKTKRALCTITLIGTFVLGALALALMLKFMGS--- 111
Db 218 RVLTL--PDAFVCR-----TICTYHPN--CLFPEFYT-----NWKIESQRNV 255
OY 112 ---KCSNSGIECDSS-----SG-----TCINP-----SNMCDGVSHCGGDEBNCVRL 151
Db 256 CLKTKTESSTPSSSTPQENTISGYSLLTKRRLPEPCHSKRIYPGVDF--GGBELN----- 308
OY 152 YGDNFLOYVSSQGRKWHVPCODMNENNGAACRDMGYKKNPYSSQGIVDGSGTS--- 208
Db 309 --VTFVKGV-----NVQDE-----TCTM-IRCOFFYISLPECKEKCKC 347
OY 209 FKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCIACGVN--LNSRHSRIVGGESALPGA 266

```

Db	348	FLRLSMDSFPRIR--AVGTGGSSG---YSRLCNTGDSNVCTKTSTRILYVCGNNSWGE	401
Qy	267	MPWQVSLHVG---NVHVCSSGIITPEMLVTPAHGVE-KPLNPPNPHMTAFAGILRQSMFY	322
Db	402	MPWQVSLVQVKTLPQRRHLCSSGLSHQWLVTAHCFDESLPDQVWR--TYSGLILSDITK	459
Qy	323	GAGY-QVEKVIASHNDYDSKTKNDIALMLKLOKPLTFNDLVKVCSLPPRGMLLPEQOLCI	381
Db	460	DTPEPSQKEIILIHQNYKASBGNHDIALLIKLQAPPLNTYEPQKRICPLSPKSDTSTLYTCWY	519
Qy	382	SGWGATEEKGTSLEYLNAKAVLLIETQRCNSRYVYDNLITPRMILCAGFLQGNVDSCGDS	441
Db	520	TGMEFSKSEKGEIQIOLIKVNIPLVTNEBCQRY-ODYKTIQPMYACAGYKKEGKADCKGDS	578
Qy	442	GGPIVTSKNNITWMLIGDTSMSGGCAKARPEVYGVNVAWFFDMYIRQMR-ADG	492
Db	579	GGPIVCKHNGMRLVGLTSGEGCARPEQVYVYKVALEIMDIIMLEKQSSDG	630

CC	RESULT	9
CC	HEPS_HUMAN	
ID	HEPS_HUMAN	STANDARD: PRT; 417 AA.
AC	P05981;	
DT	01-NOV-1988 (Rel. 09, Created)	
DT	01-NOV-1988 (Rel. 09, Last sequence update)	
DT	01-NOV-1997 (Rel. 35, Last annotation update)	
DE	SERINE PROTEASE HEPsin (EC 3.4.21.-) (TRANSMEMBRANE PROTEASE, SERINE 1).	
GN	HFN OR TMPRSS1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Liver;	
RX	MEDLINE=88209431; PubMed=28305076; Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davie E.W., "A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells."; Biochemistry 27:1067-1074(1988).	
RL	[2]	
RN	CHARACTERIZATION.	
RP	MEDLINE=91358502; PubMed=1885621; Tsuji A., Torres-Rosado A., Atal T., le Beau M.M., Lemons R.S., Chou S.H., Kurachi K.; "Hepsin, a cell membrane-associated protease. Characterization, tissue distribution, and gene localization."; J. Biol. Chem. 266:16948-16953(1991).	
RL	[3]	
RN	CHARACTERIZATION.	
RP	MEDLINE=93346237; PubMed=8346233; Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.; "Hepsin, a putative cell-surface serine protease, is required for mammalian cell growth."; Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).	
RL	-1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE OF CELL MORPHOLOGY.	
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.	
CC	-1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES, WITH THE HIGHEST LEVEL IN LIVER.	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
CC	EMBL: M18930; AAA36013.1; -.	

DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (MATRIPTASE) (MEMBRANE-
 DE TYPE SERINE PROTEASE 1) (MT-Sp1).
 GN ST14 OR PRS14 OR SRC19.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99303581; PubMed=10373424;
 RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
 RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine
 RT protease with trypsin-like activity."
 RL J Biol. Chem. 274:18231-18236(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Takeuchi T., Shuman M.A., Craik C.S.;
 RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
 RT dissect complex biological processes and identify a membrane-type
 RT serine protease in epithelial cancer and normal tissue."
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
 RN [3]
 RP CHARACTERIZATION.
 RC TISSUE=Milk;
 RX PubMed=10373425;
 RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
 RT "Purification and characterization of a complex containing matrilysin
 RT and a Kunitz-type serine protease inhibitor from human milk."
 RL J. Biol. Chem. 274:18237-18242(1999).
 CC -1- FUNCTION: DEGRADERS EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
 CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
 CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
 CC OR LYS AS THE P1 SITE.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF118224; AAD42765.2; -;
 DR EMBL: AF133086; AAF00109.1; -;
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.302; -;
 DR InterPro: IPR000859; -;
 DR InterPro: IPR001254; -;
 DR InterPro: IPR001314; -;
 DR InterPro: IPR002172; -;
 DR Pfam: PF00057; 1d1_recept_a; 4.
 DR Pfam: PF00089; Trypsin; 1.
 DR Pfam: PF00431; CUB; 2.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS00068; LDLRA_2; 4.
 DR Signal anchor: Glycoprotein; Hydrolase; Serine protease;
 KW Transmembrane; Repeat.
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 77 855 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 214 334 CUB 1.
 FT DOMAIN 340 447 CUB 2.
 FT DOMAIN 452 487 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 487 524 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 524 560 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 566 603 LDL-RECEPTOR CLASS A 4.
 FT DOMAIN 614 851 CATALYTIC.
 FT ACT_SITE 656 656 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 805 805 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;
 Query Match 20.6%; Score 558.5; DB 1; Length 855;
 Best Local Similarity 33.9%; Pred. No. 3.4e-34;
 Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;
 QY 110 GSKCSNSGIECDSSGTCINPSMNCDSVSHCPGSGEDNRCVRLYGPFFILQVYSQKRSWH 169
 DB 522 GCSCPQATERC-SNGKCLSKSQOQNGKDCGDSDEASCPKV---NVV----- 565
 QY 170 PQCQDDMNENYGRACRDMGKKNFYSSOGIYDSDGSTSPMKNTSAGVNDYKKLYHSD 229
 DB 566 -----TCT-----KHTYRLNG-----LCLKGNPECDGKDCSD 595
 QY 230 ACSKRAVSLRACIACVNLNS-SRQRIYGESALPGAMPQVSLH-VONVHVGSGIIT 287
 DB 596 GSDEK-----DCDGLRSPFRQARVVGTDADGEMPMQVSLHALQGHICGASLIS 647
 QY 288 PEVITVAARCVKRP-----LNNPMTAFACILRQSMFAGAQV---VEKVISHPYDSK 340
 DB 648 PMLVSAARCYIDDRGFRYSDPQWTAFLGLHDQS-QRSAPGQERRLRRIISHPEPND 706
 QY 341 TKKNDALAKKLOKLPFNDFKPVCLPNPGMLQPOLCISQMGATEKGTSEVLNA 400
 DB 707 TPDYDIALLELEKPAEYSSMVRPCLPDASHVPAKAIWYTMGHTQYGGICALILQNG 766
 QY 401 KVLLEITORNSRYVDNLTTPAMICAGFLQGNVDSQGDGSGPL-VYSKNIMWLIGD 458
 DB 767 EIVINQTCEN-LLPQQTTPRMCGVFLSGGVDSQGDGSGPLSSVEADGRI-PQAGV 823
 QY 459 TSWGSCAKAYRPGVYGNVAFPTDWI 484
 DB 824 VSWGDCACQARKPKGYTRLPFRDWI 849
 RESULT 11
 KAL_RAT STANDARD; PRT; 638 AA.
 ID KAL_RAT
 AC P14272;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN PK.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91129236; PubMed=1993180;
 RA Beaudine G., Rosinski-Chupin I., Mattei M.-G., Mbikay M., Chretien M.,
 RA Seidah N.G.;
 RT "Gene structure and chromosomal localization of plasma kallikrein."
 RL Biochemistry 30:1628-1635(1991).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RA MEDLINE=90091743; PubMed=2598771;
RA Seidah N.G., Ladeheim R., Mbikay M., Hamelin J., Lutfalla G.,
RA Rougeon F., Lazue C., Chretien M.;
RT "The cDNA structure of rat plasma kallikrein";
RL DNA 8:563-574(1989)
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HWM KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIa, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HWM KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M62357; AAA74563.1; -;
DR EMBL; M62358; AAA74563.1; JOINED.
DR EMBL; M62346; AAA74563.1; JOINED.
DR EMBL; M62347; AAA74563.1; JOINED.
DR EMBL; M62349; AAA74563.1; JOINED.
DR EMBL; M62350; AAA74563.1; JOINED.
DR EMBL; M62351; AAA74563.1; JOINED.
DR EMBL; M62352; AAA74563.1; JOINED.
DR EMBL; M62353; AAA74563.1; JOINED.
DR EMBL; M62354; AAA74563.1; JOINED.
DR EMBL; M62355; AAA74563.1; JOINED.
DR EMBL; M62356; AAA74563.1; JOINED.
DR EMBL; M62356; AAA74563.1; JOINED.
DR EMBL; M30282; AAA41463.1; -;
DR EMBL; M58590; AAA42069.1; -;
DR PIR; A39180; KORTPL.
DR HSSP; P00750; IRTE.
DR MEROPS; S01.212; -;
DR InterPro; IPR000177; -;
DR InterPro; IPR001254; -;
DR InterPro; IPR001314; -;
DR InterPro; IPR003014; -;
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0005; APPLEDOMAIN.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00495; APPLE; 4.
KW Hydrolyase; Serine protease; Glycoprotein; Plasma; zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 CATALYTIC.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.

FT DISULFID 21 104 BY SIMILARITY.
FT DISULFID 47 77 BY SIMILARITY.
FT DISULFID 51 57 BY SIMILARITY.
FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SO SEQUENCE 638 AA; 71273 MW; 454BB27E8CA8F98 CRC64;
Query Match 20.5%; Score 556; DB 1; Length 638;
Best local similarity 36.6%; Pred. No. 3.8e-34;
Matches 117; Conservative 59; Mismatches 128; Indels 16; Gaps 7;
QY 181 GRACRDMGYKN--NFYSOGIYVD--SGSTFMKLTSGAGNDIYKLYHSDACSSK 234
DB 314 GADACQETCTRTIRCQFTYSLPDCKAECKCSLRSTDSGSPRTIYEAGSSGYSLR 373
QY 235 AVSLRCIACGVNLNNSQSRIVGESALPGAWPROVSLHQN--HVCGSITTEPMI 291
DB 374 LCKVSSSDCTTKIN---ARIVGTSNLSGEMPOVSLQKLVSONMCGSITIGRWI 429
QY 292 VTAHCVEKPLNPNWHTAFAGILROSEPMFYGAGY-QVEKIVSPNVDYKKNNDIALMK 350
DB 430 LTAACFPD-GIPYDPVWMTYIGLNLSEITNKTPRESSIKELIHQKYMSSGSYDIALIK 488
QY 351 LQKPLTFNDLVKPYCLPMPGMMLQPEQLCWISGWGATEEKGTSEVLNAKVLLETRC 410
DB 489 LQPLNTEFEKPKICLPKADTNITVTCWVTGWTGERETONILQKATIPVPMNEC 548
QY 411 NSRYVDNLTPAMICAFELQNDSCGDSGSPVYTKNNIMWLLIGTSMGSCAKAYR 470
DB 549 OKY-RDVIYITKMCACYKKEGIDACKGDSGSLVCKHSGRWLVITSGEGCAREQ 607
QY 471 PGYGVNVAVFTDWIYQRMA 490
DB 608 PGVITKVAEYIDWILEKIQS 627
RESULT 12
ST14_MOUSE STANDARD; PRT; 902 AA.
ID ST14_MOUSE
AC P56677;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (EPITHIN).
OS ST14 OR PRSS14.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RX PubMed=10199918;
RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
RA Schwartz R.H.;
RT *Cloning and chromosomal mapping of a gene isolated from thymic
RT stromal cells encoding a new mouse type II membrane serine protease,
RT epithin, containing four LDL receptor modules and two CUB domains";
RL Immunogenetics 49:420-428(1999).

```

CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,
CC -1- AND THYMUS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART,
CC TESTIS, AND BRAIN.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF042822; AAD02230.1; -.
CC MGD: MGI:1338881; Scl14.
CC MEROPS: S01.014; -.
CC InterPro: IPR000859; -.
CC InterPro: IPR001254; -.
CC InterPro: IPR001314; -.
CC InterPro: IPR002172; -.
CC Pfam: PF00431; CUB; 2.
CC Pfam: PF00057; ldl_recept_a; 4.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00261; LDLRECEPTOR.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PROSITE: PS01209; LDLRA_1; 2.
CC PROSITE: PS50068; LDLRA_2; 4.
CC PROSITE: PS01180; CUB; 2.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC Signal-anchor: Glycoprotein; Hydrolase; Serine protease;
CC Transmembrane; Repeat.
CC TRANSMEM 1 55
CC FT 56 76 CYTOPLASMIC (POTENTIAL).
CC FT 77 902 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC FT 78 902 (POTENTIAL).
CC FT 79 902 EXTRACELLULAR (POTENTIAL).
CC FT 80 902 CUB 1.
CC FT 81 902 CUB 2.
CC FT 82 902 LDL-RECEPTOR CLASS A 1.
CC FT 83 902 LDL-RECEPTOR CLASS A 2.
CC FT 84 902 LDL-RECEPTOR CLASS A 3.
CC FT 85 902 LDL-RECEPTOR CLASS A 4.
CC FT 86 902 CATALYTIC.
CC FT 87 902 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT 88 902 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT 89 902 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT 90 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 91 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 92 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 93 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 94 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 95 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 96 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 97 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 98 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 99 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT 100 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 902 AA; 99645 MW; 2A3ED02AFA9604E CRC64;

```

Query Match 20.2% Score 548.5; DB 1; Length 902;
 Best Local Similarity 32.7%; Pred. No. 2e-33;
 Matches 123; Conservative 66; Mismatches 118; Indels 69; Gaps 13;

```

DB 594 SDGSDK-----NC-DCGLR-SFTKQARVVGTTNADEGEMPMQVSLHALGCHLGSASLI 646
QY 287 TPETWYTAHCHVEKPLANNWH-----WFAFAGILROSFMYGAGYQ---VEKYSHPNYS 339
DB 647 SDPWLVAHACHFODDKNFKYSIDYTMATFLGLDLOS-KSASGVQLKRLRIITTHSFND 705
QY 340 KTRNDIAALMLOKPLTFNDIVKPVCLPNPMDLOPEOLCWLISGWGATEKGTSEVLA 399
DB 706 FTFDYDIALLELEKSYETVVARPICLPDATHVFPKRAIWTGNCHTKEGGGALLLOK 765
QY 400 AKVLIETRCNSRYVYDNLTPAMICAGFLOGNDSCGDSGGLVYS-KNNIMWLID 458
DB 766 GEIRVINGTCED--LMPQIIFRMVCVFLSGVDSGCGDSGGLPSAEKGRMFGYV 823
QY 459 TSWGSCAKAYRPGYV 474
DB 824 VSWGEGCAQRNRPYV 839

RESULT 13
KAL_MOUSE
ID KAL_MOUSE STANDARD; PRT; 638 AA.
AC P26262;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLASMA KALLIKEIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKEIN)
DE (KININOGENIN) (FLETCHER FACTOR).
GN KLK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE=Liver;
RX MEDLINE=91090844; PubMed=2264928;
RA Seidah N.G., Sawyer N., Hamelin J., Mbikay M., Chretien M.,
RA Brachpapa L., Rochemont J., Mbikay M., Chretien M.,
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species."
RL DNA Cell Biol. 9:737-748(1990).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKEIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M58588; AAA63393.1; -.
CC PIR: A36557; KOMSPL.
CC HSSP: P00750; IRTF.
CC MEROPS: S01.212; -.
CC MGD: MGI:102849; Klk3.
CC InterPro: IPR001177; -.
CC InterPro: IPR001254; -.
CC InterPro: IPR001314; -.
CC InterPro: IPR003014; -.
CC Pfam: PF00024; PAN; 4.

```

DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00005; APPELDOMAIN.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS00495; APPEL; 4.
 KW Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 KW Repeat.
 FT SIGNAL 1 19
 FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
 FT DOMAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
 FT DOMAIN 20 105 Apple 1.
 FT DOMAIN 110 195 Apple 2.
 FT DOMAIN 200 285 Apple 3.
 FT DOMAIN 291 376 Apple 4.
 FT DOMAIN 389 621 CATALYTIC.
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 453 453 PROBABLE.
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
 FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
 FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
 FT ACT_SITE 21 104 BY SIMILARITY.
 FT DISULFID 47 77 BY SIMILARITY.
 FT DISULFID 51 57 BY SIMILARITY.
 FT DISULFID 111 194 BY SIMILARITY.
 FT DISULFID 137 166 BY SIMILARITY.
 FT DISULFID 141 147 BY SIMILARITY.
 FT DISULFID 201 284 BY SIMILARITY.
 FT DISULFID 227 256 BY SIMILARITY.
 FT DISULFID 231 237 BY SIMILARITY.
 FT DISULFID 292 375 BY SIMILARITY.
 FT DISULFID 318 347 BY SIMILARITY.
 FT DISULFID 322 328 BY SIMILARITY.
 FT DISULFID 340 345 BY SIMILARITY.
 FT DISULFID 383 503 BY SIMILARITY.
 FT DISULFID 419 435 BY SIMILARITY.
 FT DISULFID 517 584 BY SIMILARITY.
 FT DISULFID 548 563 BY SIMILARITY.
 FT DISULFID 574 602 BY SIMILARITY.
 SQ SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;

Query Match 19.8%; Score 538; DB 1; Length 638;
 Best Local Similarity 37.1%; Pred. No. 8.3e-33;
 Matches 115; Conservative 52; Mismatches 117; Indels 26; Gaps 6;

QY 185 CRDMGKNNFYSSOGIVDGSSTFMLKNTSAGNVDYKLYHSDACSSKAVALSLRCAIC 244
 DB 340 CKERGGKSLRLS---FDGSPRLITYGQSSGSLRLCKLVDPDCTTKL----- 387
 QY 245 GVNINSSNQSIVGESLPGAMPVQVSLYQVNV---HVCGSGITTPRWITAAHCVEKP 301
 DB 388 -----NAIVGCTNALSLEMPVQVSLYQVNV---HVCGSGITTPRWITAAHCVEKP 301
 QY 302 LNNPMTAFAGILROSMFYGA-GYQVEKVISHPNDYKTKNDIALMOKPLTENDL 360
 DB 439 IPRVDVRIYIGIILSLSEITETKPSRKELIHOEYKVSAGNDIALIKIQTPLNTYEF 498
 QY 361 VKPVCPLNPGMMLQPEOLCWISGKATEEKCTSEVLNAKVLLEIQRCSRYVDNLI 420
 DB 499 QKPICLPSKADNTIYNCVWTGMYTKEGEGTONIILQKATIPLVNPECCQKY-RDVI 557
 QY 421 TPANICAGFLDGNVDSGCGSGGLVYSKNNIMLIGDTSNGSCAKYRPGVGNWVF 480
 DB 558 NKOMICAGYKCGTADCKGDSGGLVYCKHSGRMQLVGTISWEGCGGRKDPGYVTVKSEY 617
 QY 481 TDVIYRQMA 490
 DB 618 MDVILEKTQS 627

RESULT 14
 HEP5_MOUSE STANDARD; PRT: 416 AA.
 ID HEP5_MOUSE
 AC O35453;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE SERINE PROTEASE HEP5IN (EC 3.4.21.-).
 GN HPN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver.
 RX MEDLINE=98058912; PubMed=9395459;
 RA Vu T.-K.H., Liu R.W., Haakma C., Tomasek J.J., Howard E.W.;
 RT "Identification and cloning of the membrane-associated serine
 protease, hepsin, from mouse preimplantation embryos."
 RL J. Biol. Chem. 272:31315-31320(1997).
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
 CC OF CELL MORPHOLOGY.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF030065; AAB84221.1; -.
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.224; -.
 DR MGD; MGI:1196620; Hpn.
 DR InterPro; IPR001254; -.
 DR InterPro; IPR001314; -.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
 FT CHAIN 1 161 NON-CATALYTIC CHAIN (POTENTIAL).
 FT CHAIN 162 416 CATALYTIC CHAIN (POTENTIAL).
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 17 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 44 416 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 162 416 CATALYTIC.
 FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 256 256 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 352 352 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 187 206 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 152 276 BY SIMILARITY.
 FT DISULFID 321 337 BY SIMILARITY.
 FT DISULFID 348 380 BY SIMILARITY.
 FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 416 AA; 44739 MW; 432194FF4004F848 CRC64;

Query Match 19.7%; Score 536.5; DB 1; Length 416;
 Best Local Similarity 30.6%; Pred. No. 6.7e-33;
 Matches 132; Conservative 63; Mismatches 180; Indels 57; Gaps 9;
 QY 77 CTSKTKALCTITLFTVIGALALAGLLMKFMGSKNSGIEDSGTINPSMCDGV 136
 DB 11 CCSRPKVAALIVGTL-LFTG-----IGASMAIVT 40

```

QY 137 SHCPGDEGENRCVRLYGPFFILQVYSSQRKSWHPYQODDMMENYGRACRDMKYNNEYS 196
DB 41 ILLQSDQDEFLYOVLSPEQSRILAVLDKTEGTRRLCQSSNARVAGLCEEGFLALAH 100
QY 197 SGGIVDDSGSTSMKLNLSA-----GNVDYIKLYHSDA---CSKAVVSLRCIACGVN 247
DB 101 SFLDYRTAGAN-----GTSGFPCVDEGGLPLAQRLILDVSYVDCPRGRRLTATCQDCG-- 153
QY 248 LNSSRQSRIVGCESSALPGAMPQVSLHQNVAHVCGSITTPPMIYTAHCVEKPLNNPMH 307
DB 154 RRKLPEYRIVGQSDSLGKMPQVSLRDGTHLCGSLISGVMVLTAAHCFERNVLSR 213
QY 308 WFAFAGILRQSMFPGAGYVEKVIKSHPMY-----DSKTKNNDIALMLQKPLPFENDLY 361
DB 214 WVEFGAVARITSP-HAVVLGVDAVVIYHGGLPFRDPTIDENSNDIALVHLSLPLETYI 272
QY 362 KPVCLPMPQMLQPOLCWLISGMGATEEKGKTSVLAIAKVLIIETORCNSRYVYDNLIT 421
DB 273 QVVCLEPAAGQALVVDKVCYVTGMNTQFYGQQAAMVLQEARVPLISNEVCNSPDFYGNQIK 332
QY 422 PAMTCAGFLQGVNDSCQDGGPLY---TSKNNTWMLIGDTSMGSGCKAKAYRPGYGVN 477
DB 333 PRMFCAGYPEGIDACQDGGSPFCVCEDSISGTSRWRLCGIIVSWGTCALARKPGYVTV 392
QY 478 WFTDMIVYROMR 489
DB 393 TDFREWIFKAIR 404

RESULT 15
FAIL_HUMAN
ID FAIL_HUMAN STANDARD; PRT; 625 AA.
AC P03951.
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COAGULATION FACTOR XI PRECURSOR (EC 3.4.21.27) (PLASMA THROMBOPLASTIN
DE ANTECEDENT) (PTA).
CN F11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86243360; PubMed=3636155;
RA Fujikawa K., Chung D.W., Hendrickson L.E., Davie E.W.;
RT "Amino acid sequence of human factor XI, a blood coagulation factor
RT with four tandem repeats that are highly homologous with plasma
RT prekallikrein."
RL Biochemistry 25:2417-2424(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88107663; PubMed=2827746;
RA Asakai R., Davie E.W., Chung D.W.;
RT "Organization of the gene for human factor XI."
RL Biochemistry 26:7221-7228(1987).
RN [3]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=91152017; PubMed=1998667;
RA McMullen B.A., Fujikawa K., Davie E.W.;
RT "Location of the disulfide bonds in human coagulation factor XI: the
RT presence of tandem apple domains."
RL Biochemistry 30:2056-2060(1991).
RN [4]
RP VARIANT LEU-301.
RX MEDLINE=90046656; PubMed=2813350;
RA Asakai R., Chung D.W., Ratnoff O.D., Davie E.W.;
RT "Factor XI (plasma thromboplastin antecedent) deficiency in Ashkenazi
RT Jews is a bleeding disorder that can result from three types of point
RT mutations."
RL Proc. Natl. Acad. Sci. U.S.A. 86:7667-7671(1989).

```

```

RN [5]
RP VARIANT LEU-301.
RX MEDLINE=92190478; PubMed=1547342;
RA Meijers J.C., Davie E.W., Chung D.W.;
RT "Expression of human blood coagulation factor XI: characterization of
RT the defect in factor XI type III deficiency."
RL Blood 79:1435-1440(1992).
CC -1- FUNCTION: FACTOR XI TRIGGERS THE MIDDLE PHASE OF THE INTRINSIC
CC PATHWAY OF BLOOD COAGULATION BY ACTIVATING FACTOR IX.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF ARG-1-ALA AND ARG-1-VAL
CC BONDS IN FACTOR IX TO FORM FACTOR IXA.
CC -1- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND. AFTER ACTIVATION
CC THE HEAVY AND LIGHT CHAINS ARE ALSO LINKED BY A DISULFIDE BOND.
CC -1- PTM: ACTIVATED BY FACTOR XIIA (OR XII), WHICH CLEAVES EACH
CC POLYPEPTIDE AFTER ARG-387 INTO THE LIGHT CHAIN, WHICH CONTAINS THE
CC ACTIVE SITE, AND THE HEAVY CHAIN, WHICH ASSOCIATES WITH HIGH
CC MOLECULAR WEIGHT (HMW) KININOGEN.
CC -1- DISEASE: DEFECTS IN F11 ARE A CAUSE OF A BLOOD COAGULATION
CC ABNORMALITY (ROSENTHAL SYNDROME) OCCURRING IN HIGH FREQUENCY IN
CC ASHKENAZI JEWS.
CC -1- SIMILARITY: CONTRAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M13142; AAA52487.1; -
DR EMBL: M20218; AAA51985.1; -
DR EMBL: M18296; AAA51985.1; JOINED.
DR EMBL: M21184; AAA51985.1; JOINED.
DR EMBL: M18298; AAA51985.1; JOINED.
DR EMBL: M18299; AAA51985.1; JOINED.
DR EMBL: M18300; AAA51985.1; JOINED.
DR EMBL: M18301; AAA51985.1; JOINED.
DR EMBL: M18302; AAA51985.1; JOINED.
DR EMBL: M18303; AAA51985.1; JOINED.
DR EMBL: M18304; AAA51985.1; JOINED.
DR EMBL: M19417; AAA51985.1; JOINED.
DR EMBL: M20217; AAA51985.1; JOINED.
DR PIR: A27431; KFH01.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.213; -.
DR MIM: 134540; -.
DR MIM: 264900; -.
DR InterPro: IPR000177; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR InterPro: IPR003014; -.
DR Pfam: PF000024; PAN; 4.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00005; APPLDOMAIN.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS00495; APPL; 4.
KW Hydrolase; Serine protease; Glycoprotein; Plasma; Blood coagulation;
KW Duplication; Signal; Disease mutation.
FT SIGNAL 1 18
FT CHAIN 19 387
FT CHAIN 388 625
FT DOMAIN 19 104
FT DOMAIN 109 194
FT DOMAIN 199 284
FT DOMAIN 230 375
FT DOMAIN 384 625
FT CARBOHYD 90 90
FT CARBOHYD 126 126

```

```
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .).
FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .).
FT ACT_SITE 431 431
FT ACT_SITE 480 480
FT ACT_SITE 575 575
FT ACT_SITE 20 103
FT DISULFID 29 29 WITH A CYS RESIDUE.
FT DISULFID 46 76
FT DISULFID 50 56
FT DISULFID 110 193
FT DISULFID 136 165
FT DISULFID 140 146
FT DISULFID 200 283
FT DISULFID 226 255
FT DISULFID 230 236
FT DISULFID 291 374
FT DISULFID 317 346
FT DISULFID 321 327
FT DISULFID 339 339
FT DISULFID 380 500 INTERCHAIN.
INTERCHAIN (BETWEEN HEAVY AND LIGHT
CHAINS).
FT DISULFID 416 432
FT DISULFID 514 581
FT DISULFID 545 560
FT DISULFID 571 599
FT VARIANT 301 301
FT CONFLICT 226 226 F -> L (IN F11 DEFICIENCY).
/FTID-VAR_006622.
SQ SEQUENCE 625 AA; 70109 MM; 147AFR94B709EBF CRC64; C -> S (IN REF. 2).
```

Query Match 19.6%; Score 533; DB 1; Length 625;
Best Local Similarity 30.5%; Pred. No. 1.9e-32;
Matches 122; Conservative 63; Mismatches 129; Indels 86; Gaps 11;

```
OY 169 HPVC-----ODDNENYGRACRDMGKNNFYSGIVDDSGSTFEMKLTISAGNDIY 222
Db 233 HPGCLFPTFESQEMPKESQRMCLKTSEGLPSTR--IKKSKALSGFSLQSCRHSIPVF 290
OY 223 --KKLYHS-----DACSSKAVSLRCL-----ACGVNL 248
Db 291 CHSSEYHDTDLGERLDIYAAKSHRACQLCTNNAVRCQFTYTPAQASCNBEGKCYLKL 350
OY 249 NSSR-----OSRIVGESALPGAMPQVSLHQN 278
Db 351 SSNGSPFKILHGRGISGYTLRLCKMNECTTKIKPRIVGTAAYRGEMPMQVTLHTTSP 410
OY 279 ---HYCGGSITTPREYIVYAAHCEKPLNNPMHMTAFACILROSFM-----FYGAGYQVEK 330
Db 411 TORHLGGSIIGNOMILTAHCF-YGVESPKILRVSGILNQSEIKEDTSFFG---VOE 465
OY 331 VISHPNYDSTKRNNDIALMKLQKPLTFNDLVKPYCLPNPGMMLQPEOLCMIWISAGNATEEK 390
Db 466 ITHIDQYMAESGYDIALKLETIVNTDSQRPICLPKSGDRNVYITDCWVTGMYRKLK 525
OY 391 GKTSEVLAAKAVLLIETQRCSRYVDNLITPAMICAGFLOGNVDSCGDSGGLVYTSKN 450
Db 526 DKIONTLQAKIKPLVTNECCQKRY-RGKHITHKMICAGYREGGKDACGDSGGLSCKN 584
OY 451 NIMWLIGDTSWSGCAKAYRPGYGVNVPFTDWIYROMRA 490
Db 585 EVMHLVGITSWGEGCAQERERPGVYTNVEYVDWILEKTOA 624
```

Search completed: September 26, 2001, 09:27:26
Job time: 316 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:45 : Search time 40.65 Seconds
(without alignments)
1601.330 Million cell updates/sec

Title: US-09-615-285-2
Perfect score: 2717
Sequence: 1 MALNCSGPAIGPYENHGY.....YIGNVWFPTDIYRQRADG 492

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_unclassified:*
14: sp_vertebrate:*
15: sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1009	37.1	767	13 Q9DGR2	Q9dgr2 xenopus lae
2	684	25.2	423	4 Q9NZAS	Q9nzas homo sapien
3	676.5	24.9	437	4 Q9NBS4	Q9nbs4 homo sapien
4	664.5	24.5	445	11 Q9ER03	Q9er03 mus musculu
5	651.5	24.0	457	4 Q9H3S3	Q9h3s3 homo sapien
6	601	22.1	1042	4 Q9Y5O5	Q9y5o5 homo sapien
7	588	21.6	311	11 Q9ER04	Q9er04 mus musculu
8	587.5	21.6	273	11 Q9ER02	Q9er02 mus musculu
9	577.5	21.3	1113	11 Q9Z3I9	Q9z3i9 mus musculu
10	564.5	20.8	855	11 Q9J1I7	Q9j1i7 rattus norv
11	563.5	20.7	643	6 Q9Y5O6	Q9y5o6 sus scrofa
12	558.5	20.6	855	4 Q9HCA3	Q9hca3 homo sapien
13	554.5	20.4	855	4 Q9H3S0	Q9h3s0 homo sapien
14	544	20.0	1379	5 Q9V4N6	Q9v4n6 drosophila
15	543	20.0	845	13 Q9DGR1	Q9dgr1 xenopus lae
16	533	19.6	571	4 Q9Y4S5	Q9y4s5 homo sapien
17	529	19.5	310	11 Q9QY29	Q9qy29 mus musculu
18	527.5	19.4	418	4 Q60235	Q60235 homo sapien
19	522.5	19.2	812	11 Q9R0W3	Q9r0w3 rattus norv

20	509.5	18.8	329	6 Q9GLI0	Q9gli0 ovis aries
21	509.5	18.8	389	13 Q9PVX7	Q9pvx7 xenopus lae
22	503.5	18.5	279	11 Q9QZ74	Q9qz74 rattus norv
23	501	18.4	1524	13 Q9I674	Q9i674 xenopus lae
24	500.5	18.4	305	11 Q9JH07	Q9jh07 mus musculu
25	499.5	18.4	868	5 Q9Y1V3	Q9y1v3 polyandroca
26	496	18.3	422	4 Q9UL52	Q9ul52 homo sapien
27	493	18.1	307	4 Q9UHW2	Q9uhw2 homo sapien
28	493	18.1	1186	5 Q9VSU2	Q9vsu2 drosophila
29	493	18.1	1449	5 Q9U1I2	Q9u1i2 drosophila
30	493	18.1	1462	5 Q9U1I3	Q9u1i3 drosophila
31	490.5	18.1	415	6 Q29015	Q29015 sus scrofa
32	489.5	18.0	581	5 Q9XZM7	Q9xzm7 strongyloce
33	473.5	17.4	421	11 Q60491	Q60491 cavia porce
34	468.5	17.3	311	11 Q9QU17	Q9qu17 mus musculu
35	468.5	17.3	312	4 Q9NS34	Q9ns34 homo sapien
36	468.5	17.2	321	4 Q9NR08	Q9nr08 homo sapien
37	468	17.2	297	11 Q88781	Q88781 rattus norv
38	467.5	17.2	321	4 Q9UBR2	Q9ubr2 homo sapien
39	466.5	17.2	806	6 Q18783	Q18783 macropus eu
40	463	17.0	273	6 Q9XSM2	Q9xsm2 ovis aries
41	462	17.0	317	4 Q9GZM4	Q9gzm4 homo sapien
42	460.5	16.9	306	11 Q9ER10	Q9er10 mus musculu
43	459.5	16.9	314	4 Q9Y6M0	Q9y6m0 homo sapien
44	459	16.9	342	11 Q9ES87	Q9es87 rattus norv
45	459	16.9	342	11 Q9ER01	Q9er01 rattus norv

ALIGNMENTS

RESULT 1
Q9DGR2 PRELIMINARY: PRT: 767 AA.
AC Q9DGR2;
ID 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE EMBRYONIC SERINE PROTEASE-2.
GN XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
from Xenopus laevis.";
RL Gene 252:209-216(2000).
DR EMBL; AB038497; BAB08217.1; -.
KW Protease.
SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;

Query Match 37.1%; Score 1009; DB 13; Length 767;
Best Local Similarity 50.9%; Pred. No. 9.9e-84;
Matches 189; Conservative 49; Mismatches 131; Indels 2; Gaps 2;

QY 120 CGSSGCIIPNSMCDGVSHSGEDENRCVRLYGNPFILOYVSSQKSMHPVCCDDMNEN 179
DB 395 CGSSVSCVLSQWCDVSCPPYGEDEMCSLTPADFOVSTVSAMLPVCSPTWDD 454
QY 180 YGRACRDMGKYNFNYS-QGIYDDSGSTSEFKMLTNSAGNVLYKKLYHSDACSSKAVS 238
DB 455 FGRFACQDGRGVSSGNRDTLMSPYAPNGYFKIYSGYRSKFTSYGVYSSVCYGNVYS 514
QY 239 LRFACGVNLSSRSRITVGGESALPGAMPQVSLHYOVNHYVGGSIITPEYITVAACHV 298
DB 515 LHCISGVSNNSLVSRIYGTGFANIGNMPQVNLQYITGVLCGGSITSPKWIYVAACHV 573

QY	299	EKPLPMHWAFFAGILDSQFMEFAGQVQKXVSHNNYSKTNNDIALMLKLOKPLETN	358
		::: : : : : : : : : : :	
Db	574	YGSYSSASGMRVFGTLTKPSYNASAFVBRLLVHFGYTSYTDNDIALMLKLEDTFG	633
QY	359	DLYKPVCLPNFGMMLQPEOLCMISMGATEBKGTSEYLANAKVLLIETORCSNRYVDN	418
		: : : : : :	
Db	634	YTTQPVCLPNSGMFMWEGATTWISMGSTYEGGSVSYLYQAAFLPDSMVCNQSYYNC	693
QY	419	LITPAMICAGLOGNDSCGDSGCLPVTSKNNITWMLIGTSMGSCGAKAKRPGVYGNM	478
		: : : : : : : :	
Db	694	QITSMITACAGLSSGVDPTCQDSGGLPVNKRNGTWMLVGDTSWGDGCAKANKPGVYNT	753
QY	479	VFTDMYIROMR	489
		:	
Db	754	TFLMYSQMR	764

```

09NZAS      2
RESULTZS    2
ID          09NZA5 PRELIMINARY; PRT;   423 AA.
AC          09NZA5:
DT         01-OCT-2000 (TREMBLE). 15. Created)
DY         01-Oct-2000 (TREMBLE). 15. Last sequence update)
DT        01-MAR-2001 (TREMBLE). 16. Last annotation update)
DE TYPE II MEMBRANE SERINE PROTEASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Plimates; Carnivora; Homnidae; Homo.
RN NCBI_Taxid=9606; [1]
RP SEQUENCE FROM N.A.
RA Smeekens S.S., Lotimer D.D., Wang E., Hou J., Linnevers C.;
RT "Wt-Sp2, a novel type II membrane serine protease expressed in
RT tracheas, colon, and small intestine: Identification, cloning, and
RT chromosomal localization."
CC Submitted (DEC-1999) to the EMBL/GenBank/DDBB databases.
RI - SIMILARTY: TO CHYTOTRYPSIN SERINE PROTEASE FAMILY (SI).
DR EMBL_AFP1612; AAF31436.1; -.
DR InterPro: IPR001284; -.
DR InterPro: IPR001374; -.
DR InterPro: IPR002172; -.
DR Pfam: PF00057; Idl_recept.a; 1.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR PROSITE: PS00134; TRypSin_HIS; UNKNOWN_1..
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR SMART: SM00192; IDLa; 1.
KV Protease.
SQ SEQUENCE       423 AA: 46387 MW; 90792AFOF8AFEAB30 CRC64;
```

Query Match	25.20;	Score 684;	DB 4;	Length 423;
Best Local Similarity	36.20;	Pred. No. 2.5e-54;		
Matches 165; Conservative	64;	Mismatches 167;	Indels 60;	Gaps 16;

```
OY      61 SNPVACGQPKSPSTVCTSKTKKALCITLTGTGLVGAALAAAGLTFMGSCKNSGIEC 120
        ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      2 SNP--CANPVP-WRPSESGIPIIALMLSLASIIIVVLVIKLDKXY-----FLC 50

OY      121 DSSGTCINPSMWCGVSHCGEGEDENNCCVRLY--GP-----NFILQVYSQRKSMH 169
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      51 GQPLHLFIPRKOLCGELDCPLGEDDEEHCVASFPGPAVAVRLSKDNRSTLOLVSATATGNMF 110
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

OY      170 PVCODDNNENYGRAACHDMGY--KNNEYSSQ-----GIYDDSGSIFEMKLNTSGAND 220
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      111 SACDNFTFEALAEACQOMGYSKPFPEAAVEIGPDOLDVAITEITENSQELRMNNSG--- 167
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

OY      221 IYRKLTHSDACSSKAAYVSLGCIAGCVGLNLSROSRIRVGESALPGAMPQVSVLHYGVNH 280
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      168 -----PCLSLSLVSLNCLTACGKSLL----KTPRVVGGEEAASVDSPMVVSITYDKQH 215
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

OY      281 CGGSITPEWIVTLAHACEVKERPLNNPMHWTAFACILR-QSFMEFYAGCYOVEKVY---SHPN 336
        ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

[illegible]

RESULT	3			
ID	Q9NR54	PRELIMINARY;	PRT;	437 AA.
AC	Q9NR54;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-MAR-2001 (TrEMBLrel. 15, Last annotation update)			
DE	TRANSMEMBRANE SERINE PROTEASE 3.			
GN	TPRSS3.			
OS	Scaplaens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Euteleia; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=PANCREATIC CARCINOMA;			
RA	Medline-20283276; PubMed-10825129;			
RA	Wallrapp C., Hahnle S., Muller-Pillasch F., Burghardt B., Iwamura T.,			
RA	Ruttenburger M., Lerch M.M., Adler G., Gress T.M.,			
RT	"A novel transmembrane serine protease (TPRSS3) overexpressed in			
RT	pancreatic cancer."			
RL	Cancer Res. 60:2602-2606(2000).			
CC	-1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (SI).			
DR	EMBL: AF179224; AAF74526.1; "			
DR	InterPro: IPR001254; "			
DR	InterPro: IPR001314; "			
DR	InterPro: IPR002172; "			
DR	Pfam: PF00057; Id1_recept_a; 1.			
DR	Pfam: PF00089; trypsin; 1.			
DR	PRINTS: PR00722; CHYMOTRYPSIN.			
DR	PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.			
DR	PROSITE: PS00135; TRYPSIN_SER; 1.			
DR	SMART: SM00192; Ldla; 1.			
KW	Protease.			
SC	SEQUENCE 437 AA; 48204 MW; 351B2FDDA8657B12 CRC64;			

Query Match	24.9%;	Score 676.5;	DB 4;	Length 437;
Best Local Similarity	39.1%;	Pred. No. 1.3e-53;		
Matches 150;	Conservative 57;	Mismatches 128;	Indels 49;	Gaps 13;

Qy	133	CDGVSHPGGDEENRCVRLY--GG-----NFIIQVYSQRRSHNRCPODDMENEY	181
Db	77	CDGELDRLPGGDEENHCYKSPFEGRAVAVALRLSKDSTLDVLDSATGNNFSCFDNFTAL	136
Qy	182	RAACRDGMGY--KNNFYSSQ-----GIYDSSGSTSPMKLNTSAGNVLYKLYHSDAC	232
Db	137	ETACRQMGYSKPFPRFAVEIGDPDLDLVLETENSSQLRRNRSSG-----PCL	184
Qy	233	SKAVYSLRLCIACGVNLLMSRQSRIVYGEESALPGAMPQVQSLYHQNVHVCGSITTEPMY	292
Db	185	SGSLVSLHLCIACGKSL--KTPRVAVGGEESVDSWPQVQSIQYDKQVCGSITLDPWYL	241
Qy	293	TAALHCEVRPLNNHPTAFAGILR-QSFMYGAGUYQEVKY---SHPNYDSKTRNDIAL	348
Db	242	TAALHCFKHN-IDVFNKMKVLRAGSDKLTGSF-----PSLAVALKIIITIEFNMY---PKNDIAL	293
Qy	349	MKLOKPLTFENDLVKRVSLPFGKMLDREOLCWSIGMGATEEK-GKTSFVYLAARVLIET	407

Db 294 MRLQPLFFSGTVRRPCLPFDEELTPATPLWIGMGFKONGKMSDILLQASVOIDS 353
 Oy 408 QRCNRRYVDNLITPRAMICAGFLQGNVDSGGSGPLVTSKNNIMWLIGDTSWSSGCK 467
 Db 354 TRCNADDAVQGEVTEKMCACIPEGGVDICQDSSGGLVYOSDO-WHVVGIYSWVGCGG 412
 Oy 468 AYRPGVGNVWFDTMIYFROMRAD 491
 Db 413 PSTPGVTKVSAVLMWITYNWKA 436

RESULT 4

Q9ER03 PRELIMINARY: PRT: 445 AA.

AC 09ER03
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE TYPE 3 SPINESIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN:
 RA Mitsui S., Yamaguchi N.;
 RT "cDNA cloning of mouse spinesin."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016230; BAB20277.1; -
 SQ SEQUENCE 445 AA; 48502 MW; F60E7CEA6567C7E7 CRC64;

Query Match 24.5%; Score 664.5; DB 11; Length 445;
 Best Local Similarity 35.0%; Pred. No. 1.6e-52;
 Matches 145; Conservative 66; Mismatches 176; Indels 27; Gaps 9;

Db 86 CTTLGLTFVGAALAGLIMKFKMSKCSNSGIECDSTGTCINPNCWCDGVSHCGEDE 145
 Oy 146 NRCV-----RLYGNFTLQVYSSQKSWHPYCODMNNENYGRACRDMGY-KNNFY 195
 Db 90 BELRLSLKTVSFRINGEDLLQVQVRRPDLVLCHEGWSFALMHNICKSGHLRLQH 149
 Oy 136 SSGQIVDD--SGSISFMKLNYSAGVNDIYKLYHSDACSKAVLSLCTACGVNINSRQ 253
 Db 150 KAVNLSIDIKLNSQEFADLSARPGCL-VEEAWKPSANCPSCGRIVSLKSCCGAR--PLA 205
 Oy 254 SRIVGESALPGAMPVOYSLHQNHYVCGSITPEMITYTAHCVEK-PLNPNMHTAFA 312
 Db 206 SRIVGGQAVASCRPMQASVMLGSRHGTGASVLAHPWVTAHNCISFRLSLSSMRVHA 265
 Oy 313 GLRQSFMEYGAQOVYVISHPNYDSKTKNNDIALMQLKQFLTENDLYKPYCLPFGMA 372
 Db 266 GLVSHGAVRQHGTVWEKTIIPRLVSAQNHQDVALLDQRPTRINSIDVYGAICLPAKEQY 325
 Oy 373 LQPEOLCMTSGKATE-EKGTSEVLNAKVLITQRCNSRYVDNLITPRAMICAGFLQ 431
 Db 326 PPMWQSCVWVGSHDPSTHSSDPLQDTMVLSTHLCNSCMYSGALTTHMLACAGYLD 385
 Oy 432 GNVDSGCGDSGGPLVTSKNNIMWLIGDTSWSSGCAKAVRPGVYGVNMVFTMIY 485
 Db 386 GRADCGQDSSGGLVCPGSDTWHLVGVYSWGRGCAEPNRPVYAKVAEFLDWIH 439

RESULT 5
 Q9H3S3 PRELIMINARY: PRT: 457 AA.
 AC 09H3S3
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE SPINESIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN PROSTATE;
 RA Mitsui S., Yamaguchi N.;
 RT "Molecular cloning of human spinesin."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB028140; BAB20375.1; -
 SQ SEQUENCE 457 AA; 49574 MW; 64406AB4985A2651 CRC64;

Query Match 24.0%; Score 651.5; DB 4; Length 457;
 Best Local Similarity 32.1%; Pred. No. 2.6e-51;
 Matches 161; Conservative 63; Mismatches 194; Indels 83; Gaps 16;

Oy 8 PALIGPYENHGYOPENRYPAQPTVPTVYEVNRA-QYRSPVPQYARVYLQASNPVVC 66
 Db 9 PPMQAYAE-----EGPQG-----IFRAPDQDHP-----ISQA----- 39
 Oy 67 TQPKSPSGVCTSKTKKALCITLTATFLVGAALAGLIMKFKMSKCSN--SGIEDSSG 124
 Db 40 -----VCRSMRRCGAVLGAIG-LLAGAGVGSWLVLYLCRAAPISGTLQDEI 89
 Oy 125 TCINPSNMCDGVSHCPGEGDENRCV-----RLYGNFTLQVYSSQKSWHPYCODD 175
 Db 90 TL-----SCSEASAEFALLPALPKTVSFRINSEDFLEAQVRDQPMILLVCHEG 138
 Oy 176 WNEVNGRACRDMGY-KNNFYSSQGIYVDSGSGSTFEMKLNYSAGVNDIYKYL--YNSDA-- 230
 Db 139 WSPALGLQICWSLGHRLTHNKGVLTD-----IKLNSQEFADLSARPGCLVEEAWK 191
 Oy 231 -----CSSKAVLSLCTACGVNLSNSRQSRIVGESALPGAMPVOYSLHQNHYVCGSIT 286
 Db 192 PNNCTSGGVVSLKSCCGAR---PLASRIYGGQSAVAPRWQASVALGPHHTGGSVTL 248
 Oy 287 TREWITYTAHCVEK-PLNPNMHTAFAGILROSFMFYGAGYEVYISHPNYDSKTKND 345
 Db 249 APRWVYTAHCNHSFRLATLSSMRVHAGLVSHSAVPRHOGALVERIIPRLYSQNHVD 308
 Oy 346 TALMKLQKFLTENDLYKPYCLPNCMMLOPEOLCMTSGGATGAEKCK-TSEVLNAKVL 404
 Db 309 VALRLQALNFSIDVGAACLPRAKQDHPKGSRCWVGSHNPSHTYSSDMLQDTVPL 368
 Oy 405 IETORCNSRYVDNLITPRAMICAGFLQGNVDSGGSGPLVTSKNNIMWLIGDTSWSSG 464
 Db 369 FSTQLCNSSCVSGLTPMLCAGYLDGRADACQDSSGGLVCPGSDTWHLVGVYSWGRG 428
 Oy 465 CAKAYRPGVYGVNMVFTMIY 485
 Db 429 CAEPNHPGVYAKVAEFLDWIH 449

RESULT 6
 Q9Y505 PRELIMINARY: PRT: 1042 AA.
 AC 09Y505
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CORIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEART;
 MEDLINE=99262646; PubMed=10329693;


```

Db 577 GCLKKGNEPCDGGKDCSGSDEKNC-----DCGLRSFK----- 611
Oy 229 DACSSKAVYSLRCLACGVNLNSRSRIYGESALPGAMPQVSLH-VONVHVGGSIT 287
Db 612 -----QARYVCGTNADGEMPMQVSLHALGGLCAGSLIS 647
Oy 288 PEWITAAHCVEK-----PLNNPMHTAFAGILROSFMFYAGCYO--VEKVISHPNYDSK 340
Db 648 PDMVLSAHCFOEDETIFKYSDDHTMTAFGLGLDQS-KRSASGVGEHKLKRLITTHPSFND 706
Oy 341 TKNNDIAMLKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWTISGKATEBKTSFVLNAA 400
Db 707 TFDIDIALELEKPAEYSTVAPICLPDTHVFPAGKAIWVGHTREGTGALLILKG 766
Oy 401 KVLLETORCSRYVDNLITPAMICAGFLQGVNDSOGSGGLPVT--SKNNIMWLIGPT 459
Db 767 EIRVNOTTCER--LLPQITPRAMCVGFLSGGVSCGDSGGLSSVEKDRITFQAGV 824
Oy 460 SWGSCAKAYRPGVYGNVWFTDWIYRQ 487
Db 825 SWGEGCAQRNKGYYTRIPEVRDRIKEQ 852

```

```

RESULT 11
ID 097506 PRELIMINARY; PRT; 643 AA.
AC 097506;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE KALLIREIN.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLASMA;
RA Takahashi T., Kimura A., Okimura H., Hamabata T.;
RT "Porcine liver plasma kallikrein.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL; AB022425; BAA37147.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.212; -.
DR InterPro; IPR000177; -.
DR InterPro; IPR001254; -.
DR InterPro; IPR001314; -.
DR InterPro; IPR003014; -.
DR Pfam; PF00089; Trypsin; 4.
DR Pfam; PF00024; PAN; 4.
DR PRINTS; PR00005; APPELOMAMIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00495; APPEL; 4.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
DR SMART; SM00020; TRYP_SPC; 1.
DR KMW; S00000; Serine protease.
SQ SEQUENCE 643 AA; 72227 MW; AFF2923E3C3CB80A CRC64;

```

```

Query Match 20.7% Score 563.5; DB 6; Length 643;
Best Local Similarity 29.9% Pred. No. 4.5e-43;
Matches 154; Conservative 63; Mismatches 185; Indels 113; Gaps 19;
Oy 52 YAPRYLQASNPVYCTQKSPSGVYCTSKTKALCILTLGLTFYGAALAGLMLKPKMS 111
Db 160 YATQAFNNAEVRNNCLLKHSPGCTPTSIK-----VLAVNESG---FSLK 200
Oy 112 KCSNSGIECDSS-----GTCINPSMKGCGVSHCPGGEDENCVRLYGPNTLQ 159

```

```

Db 201 PCADSEICHHMIDIFOHLAFSDVDVARIAPDAFV-----C-----RTICTYHPNCLFF 248
Oy 160 VY-----SSQR-----KSMH-----PVCODDNNENYGRAACRDM---GYKNFYSSQ 198
Db 249 TFEYTNAMKIEGSRNNCFIKTSHSGTSPFTPEMNAISGTSLLTCQTLPECHSKSYIEV 308
Oy 199 GIYVDSGTSFEM-----KLNTSAGNVDIYKKLYHSAC--SSKAVVSLR----- 240
Db 309 DFEGBELNVTFOVAGANLCOEICTKTRICOFFTYSLHPEDCGKEKCKSLRSSDGPYKI 368
Oy 241 -----C-----IACGVNLNSRSRIYGESALPGAMPQVSLH-VONVHVGGSIT 287
Db 369 THGRASSGYSLRLDRSGDHSACATKAN---TRIVGTDLSLGPMPQVSLQAKLRAON 424
Oy 278 VAVCGSITPEWITAAHCVEKPLNNPMHTAFAGILROSFMFYAGCYO-OVEKVISHPN 336
Db 425 -HLCGSIIGHQWVLTAAHCD-GSLPDIRIYIGLINISEITKEMFEFSQVKEIITHQN 482
Oy 337 YDSKTKNNDIALMKLOKPLTFNDLVKPVCLPMPGMLOPEOLCWTISGKATEBKTSFV 396
Db 483 KYLESCHDIALKLETPNLNTDFEQPICLPSRDDTNVYVYNCWVGFTGFEKEGEIONI 542
Oy 397 LNAKVLLETORCSRYVDNLITPAMICAGFLQGVNDSOGSGGLPVT--SKNNIMWLIGPT 459
Db 543 LQKVNIPLVSNBECQKST-RDKKISKQMICAGYKGGKDGSGGLVCKYKNGIMHLV 601
Oy 457 GDTWSGSCAKAYRPGVYGNVWFTDWIYRQMRAD 491
Db 602 GTTWSGEGCARREGPGVYTVKIEYMDILKRTQDD 636

```

```

RESULT 12
ID 09HCA3 PRELIMINARY; PRT; 855 AA.
AC 09HCA3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SERINE PROTEASE TADG15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,
RA O'Brien T.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF057145; AAG15395.1; -.
DR KMW; S00000; Serine protease.
SQ SEQUENCE 855 AA; 94700 MW; 08827280FA2D2B3 CRC64;

```

```

Query Match 20.6% Score 558.5; DB 4; Length 855;
Best Local Similarity 33.9% Pred. No. 1.8e-42;
Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;
Oy 110 GSKCSNGIEDSSGTCINPSNMCDGVSHCPGGEDENCVRLYGPNTLQYSSQKSMH 169
Db 522 GCSCPAQTFRC-SNKKCLSKSGQOCCKGDKDGDGSEACRPV---NV----- 565
Oy 170 PVCQDDMNENTGRAACRDMGYKNFYSSQGIYVDSGTSFEMKLNTSAGNVDIYKKLHSD 229
Db 566 -----TCT-----KHTYRCLNG-----LCLSKNGNEPCDCKEPCSD 595
Oy 230 ACSSKAVYSLRCLACGVNLNSRSRIYGESALPGAMPQVSLH-VONVHVGGSIT 287
Db 596 GSDEK-----DCDGLRSFKTRQARVGGTDADEGEMPMQVSLHALGGLCAGSLIS 647
Oy 288 PEWITAAHCVEK-----PLNNPMHTAFAGILROSFMFYAGCYO--VEKVISHPNYDSK 340
Db 648 PDMVLSAHCFTIDRGRFYSDPTQMTAFGLGLDQS-QRSAGGVGERRLKRLITTHPSFND 706

```

RESULT 14	
Q9V4N6	
ID Q9V4N6	PRELIMINARY;
AC Q9V4N6;	PRT; 1379 AA

DT 01-MAY-2000 (TREMBLERel. 13, Created)
 DT 01-MAY-2000 (TREMBLERel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLERel. 16, Last annotation update)
 DE CG2105 PROTEIN.
 GN CG2105.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyndroidae; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC SPRAIN-BERKELEY.
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Aghayani A., An H.-J., Andrews-Plankkooch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Bernick B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kension J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kraevitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman K.S., Pan S., Pollard J., Pui Y., Reese M.G.,
 RA Reinett K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Stimpson M., Stropki M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Skung R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -I- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
 CC BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
 CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN (BY
 CC SIMILARITY).
 CC -I- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 CC (LDLR) DOMAIN.
 CC -I- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
 CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -I- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (SI).
 DR EMBL: AE003840; AAF59230.1; -;
 DR HSPD: P00763; IDPO.
 DR FlyBase: FBgn0033192; CG2105.
 DR InterPro: IPR000024; -;
 DR InterPro: IPR000217; -;
 DR InterPro: IPR001190; -;
 DR InterPro: IPR001254; -;
 DR InterPro: IPR001314; -;
 DR InterPro: IPR002172; -;
 DR Pfam: PF000057; Idl_recept_a; 2.
 DR Pfam: PF000089; trypsin_1.
 DR Pfam: PF01392; Fz; 1.

DR PRINTS: PR00722: CHYMOTRYPSIN.
DR PRINTS: PR00261: LDLRECEPTOR.
DR PROSITE: PS01209: LDLRA_1; 1.
DR PROSITE: PS50068: LDLRA_2; 2.
DR PROSITE: PS50287: SRCR_2; 1.
DR PROSITE: PS00134: TRYPSIN_HIS; 1.
DR PROSITE: PS00135: TRYPSIN_SER; 1.
DR PROSITE: PS00227: TUBULIN; 1.
DR SMART: SM00020: TRYP_SPC; 1.
KM GRP-binding: Glycoprotein; Hydrolyase; Microtubules; Serine protease.
SQ SEQUENCE 1379 AA; 149489 MW; A593A9CC2167EAB7 CRC64;

Query Match 20.0%; Score 544; DB 5; Length 1379;
Best Local Similarity 33.4%; Pred. No. 7e-41;
Matches 141; Conservative 65; Mismatches 158; Indels 58; Gaps 20;

QY 111 SKCSNSG---IECDSSGTCINSMNCDGSHCPGGEDENRCLYGPN-----FILOYVS 162
DB 922 AKCEGCGPEITCGDS-QCIGTKHICDGIIDCPYCODERNCLSRNEDVGTGVLEVIR 980
QY 163 SQRKSMHPVCODDMNENYGRACRDMGYKNFYSSOGIVDDSGSTSEFK-----LNT 214
DB 981 IGORQMPACVKNMORAVSPASVCSILGY-----SAVNATSVLTQLTHRPLAT 1029
QY 215 SAGNDIYK-----KLYSDACSSKA-----VSLRC--IACG-VNLNSSROS-RIVGG-259
DB 1030 VNVSTDIKMKVAKRRKSTLMQEFANCKTEDYPMADLTCSNYECGRKRRKPSRRIIG 1089
QY 260 ESALGAMPQVSL--HYQNVHVGSGSITPEWITVAACHV-EKPLNNPMTAFAGILR 316
DB 1090 TQASGNNPFLAIIIGEPKITYCAGVILSDQWVLTASICVNTSYVIDEDMTIQLGYR 1149
QY 317 Q-SFMEFYAGYQVEKYISHPNYD-SKTKNDIALMLKLOKPLTFNDLVKPVCLPNEGML-L 373
DB 1150 RNSFYSGQKVKVKAIVPHQYNNMAIADNDIALFOLATRAVFHEHLLPVCLPPSVRL 1209
QY 374 QPEQLCMISGKATSEKTKTSF---VLNAKAVLLETQRCNSRYVDNL-TTPAMICAGF 429
DB 1210 HPGTLCITYIGWCKREDKDKPKSTYEYIVNEVOPIITRNQCD--WLDNLTYSSEGVACGF 1267
QY 430 LOGNVDSGCGDSGPLY--TSKNIMWLIQDTSWGSAGCAKAYRPGYGVNVWFTDMYR 486
DB 1268 DDGADACGDSGGLPCRPYBEKRMVGVGIVSMGIMCAHRLPGVIANVOYIPWIOE 1327
QY 487 QM 488
DB 1328 QI 1329

RESULT 15
Q9DGR1 PRELIMINARY; PRT; 845 AA.
AC Q9DGR1;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE HOMOLOG OF HUMAN MT-SPL.
GN XMT-SPL.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT Isolation and characterization of three novel serine protease genes
from Xenopus laevis.
RL Gene 252:209-216(2000).
DR EMBL: AB038498; BAB08218.1;
SQ SEQUENCE 845 AA; 93597 MW; 7FD7E62851A758B7 CRC64;

Query Match 20.0%; Score 543; DB 13; Length 845;
Best Local Similarity 27.9%; Pred. No. 4.7e-41;
Matches 136; Conservative 72; Mismatches 173; Indels 106; Gaps 12;

QY 20 YQPPNPYPAQPIVPTVYEVHQAQYPPSPVQYARVLTQASNPVVCYOPSPSCTVCT 79
DB 435 YEPNRNCPDQ-----FTCRSGRCIRIDOKCDGNMDED 467
QY 80 KTKKALCITLTGTFYGAALA-----AGLLMKFPGSKNSGTECDSSGTCI 127
DB 466 FSDMSCTCTALQFCVNSKLCRPSYFICDGVNDCGSSDELACCPNNTFC-GNGKCI 526
QY 128 NPSNCDGVSHPGGEDENRCLYGPNFILQVYSSQRKSMHPVCODDMNENYGRACRD 187
DB 527 PDSQCDRYDNCGDSDEAECDQVL-----TTACTE 557
QY 188 MKYKNPFYSSOGIVDDSGSTSPMKLNTSAGNDIYKLYHSDA---GSKAVVSLRCIAC 244
DB 558 YTYK-----CKNNOCITRKKNPECDGENDCSGSDENAKCNC 594
QY 245 GVNLSRSRIRIVGESALPGAMPQVSLHYQ-NVHVGSGSITPEWITVAACHVEKP-- 301
DB 595 G-KRPFTKSRIVGVNADTGEFPMQVSLHAKGNHTCGASLGFTMLISAHCFQDDHQ 653
QY 302 --LNNPMTAFAGILROSFMFYG--AGYQVEKYISHPNYDSTKKNNDIALMLKLOKPLTF 357
DB 654 MRYSDASLMTAYLGLHDQALNTKDVVERIRIKRIAHIGFNDNTYDNDIAVLELEKPYEY 713
QY 358 NDLYKPVCLPNEGMLQPEQLCMISGKATSEKTKTSFVLNAKAVLLETQRCNSRYVD 417
DB 714 TDEIQVCIPESTHDPVYKGLPIWYTGALKEGGAVALILKAEIRIINOTECNK--LLD 771
QY 418 NLITPAMICAGFLOGNVDSGCGDSGPLYTSK-NNIMWLIQDTSWGSAGCAKAYRPGYGN 476
DB 772 GQLTPMILCAGVSGSIDACGDSGGLPSSVELNNKYVLIAGVSMGBCACARNRPGVTK 831
QY 477 VNVFTDW 483
DB 832 VSMRDM 838

Search completed: September 26, 2001, 09:27:03
Job time: 318 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: September 26, 2001, 09:22:49 : Search time 34.76 Seconds
(without alignments)
415.089 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492
Perfect score: 1319
Sequence: 1 RIVGESALPGAMPQVSLH.....VYGNVWFETDWYRQMRADG 238

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_0601:*
2: /SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT:*
3: /SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT:*
4: /SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT:*
5: /SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT:*
6: /SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT:*
7: /SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT:*
8: /SIDSR/gcgdata/geneseq/geneseq/AA1986.DAT:*
9: /SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT:*
10: /SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT:*
11: /SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT:*
12: /SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT:*
13: /SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT:*
14: /SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT:*
15: /SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT:*
16: /SIDSR/gcgdata/geneseq/geneseq/AA1994.DAT:*
17: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT:*
18: /SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT:*
19: /SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT:*
20: /SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT:*
21: /SIDSR/gcgdata/geneseq/geneseq/AA1999.DAT:*
22: /SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT:*
23: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	492	AAV92050	HRPc6/7 polypeptl
2	1319	100.0	492	AAV77726	Human tumour suppt
3	1319	100.0	492	AAV4406	Human 20P1F12-CTC2
4	1310	99.3	283	AAV81492	Human prostate-ass
5	1303	98.8	492	AAAB36901	Human TMPRSS2 prot
6	1303	98.8	492	AAV57280	Ovrl15 homolog pro
7	655	49.7	452	AAV41694	Human PRO382 prote
8	655	49.7	454	AAAB32246	Tumour associated
9	654	49.6	453	AAAB44250	Human PRO382 (UNQ3
10	645.5	48.9	248	AAAB3572	Human cancer assoc
11	645.5	48.9	327	AAV72093	Human serine prote

12	564	42.8	414	21	AAAB08912	Human secreted pro
13	564	42.8	480	21	AAAB08950	Human secreted pro
14	558	42.3	296	21	AAV72108	Human serine prote
15	558	42.3	372	21	AAV72092	Human serine prote
16	558	42.3	457	21	AAAB1699	Human serine prote
17	549	41.6	273	21	AAAB1696	Mouse serine prote
18	549	41.6	311	21	AAAB1697	Mouse serine prote
19	549	41.6	445	21	AAAB1698	Mouse serine prote
20	544	41.2	238	21	AAAB1695	Mouse serine prote
21	523.5	39.7	798	15	AAV7283	Bovine enterokinase
22	521	39.5	432	21	AAV99417	Human PRO1570 (UNQ
23	521	39.5	432	22	AAAB6581	Human PRO1570. Ho
24	521	39.5	432	22	AAAB6581	Protein of the inv
25	521	39.5	435	20	AAV06437	Human protease HUP
26	521	39.5	435	22	AAV72558	Human seripancrin
27	518.5	39.3	418	17	AAAB9435	Trypsin-like enzyme
28	518.5	39.3	418	20	AAV29498	Human lung tumour
29	518.5	39.3	418	20	AAV29501	Human lung tumour
30	518.5	39.3	418	20	AAV29502	Human lung tumour
31	518.5	39.3	418	21	AAAB4428	Human lung tumour
32	518.5	39.3	418	21	AAAB4437	Human lung tumour
33	518.5	39.3	418	21	AAAB4438	Human lung tumour
34	516	39.1	233	18	AAW22986	Human seripancrin
35	516	39.1	492	22	AAV72559	Trypsin-like enzyme
36	513.5	38.9	232	17	AAAB9430	Mouse hepsin prote
37	507.5	38.5	416	20	AAV63325	A mouse serine prote
38	507.5	38.5	416	20	AAV6812	Complete mouse pro
39	502.5	38.1	812	16	AAAB3959	Murine plasminogen
40	502.5	38.1	812	17	AAW07585	Murine plasminogen
41	502.5	38.1	812	20	AAV08686	Murine plasminogen
42	502.5	38.1	812	20	AAV94036	Murine plasminogen
43	502.5	38.1	812	21	AAAB0053	Murine plasminogen
44	502.5	38.1	812	21	AAAB16490	Murine plasminogen
45	502	38.1	683	21	AAAB19551	Human matrilysin

ALIGNMENTS

RESULT 1	
AAV92050	AAV92050 standard; Protein: 492 AA.
XX	
AC	AAV92050;
XX	
DT	01-AUG-2000 (first entry)
XX	
DE	HRPc6/7 polypeptide from androgen-inducible gene clone.
XX	
KW	Androgen inducible; testosterone; prostate cancer; cytostatic;
XX	
OS	TMRSS2; diagnosis.
XX	
OS	Homo sapiens.
XX	
PN	WO200018961-A2.
XX	
PD	06-APR-2000.
XX	
PF	30-SEP-1999; 99WO-US22535.
XX	
PR	30-SEP-1998; 98US-0163759.
XX	
PR	30-SEP-1998; 98US-0164159.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Macbeth KJ, Shyjan AW;
XX	
DR	WPI; 2000-293182/25.
XX	
DR	N-PSDB; AAA08603.
XX	
PT	Novel methods for identifying compounds for treating prostate cancer
PT	comprising measuring the level of expression or activity of 1 or more
PT	of 11 genes or their products

```
XX PS Claim 2; Fig 3; 108bp; English.
XX
CC This protein is encoded by a gene which is androgen (e.g. testosterone)
CC inducible in androgen-dependent prostate cancer cells (e.g. LNCap cells)
CC and constitutively expressed in androgen-independent prostate cancer
CC cells (e.g. LNCap cells). Agents which decrease the expression or
CC activity of these clones may slow or arrest the growth of prostate cancer
CC cells or may kill them. HRPca6/7 can be obtained from the sequence of
CC the known gene for TMPRSS2. A compound useful for treating prostate
CC cancer can be identified in a novel method comprising measuring the
CC expression level, or activity, of HRPca2, 3, 6/7, 8, 9, 10, 13, 14, 15,
CC 19, or peripheral-type benzodiazepine receptor (PBR) in a cell, in the
CC presence and absence of a test compound. The sequences may also be used
CC in diagnosis of prostate cancer and to determine efficacy of treatment
CC for prostate cancer.
XX
SQ Sequence 492 AA;

Query Match          100.0%; Score 1319; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.4e-117;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPWOVSLHVQNVHVCGSITTPEMWITAAHCVEKPLNPMHTAFAGI 60
DB 255 RIVGESALPGAMPWOVSLHVQNVHVCGSITTPEMWITAAHCVEKPLNPMHTAFAGI 314
OY 61 LRQSFMYGAGYQVEKYISHPNYDSKTRKNDIALMKLOKPLTFNDLYKPVCLPMPGMQLQ 120
DB 315 LRQSFMYGAGYQVEKYISHPNYDSKTRKNDIALMKLOKPLTFNDLYKPVCLPMPGMQLQ 374
OY 121 PEOLCWISMGATEEKGTEKTSVELNAKVLLETQRCNSRYVYDNLITPAMICAGFLQGNV 180
DB 375 PEOLCWISMGATEEKGTEKTSVELNAKVLLETQRCNSRYVYDNLITPAMICAGFLQGNV 434
OY 181 DSCGDSGGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGVNMFDMITRQMRADG 238
DB 435 DSCGDSGGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGVNMFDMITRQMRADG 492

RESULT 2
AAV77726
ID AAV77726 standard; Protein: 492 AA.
XX
AC AAV77726;
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMPRSS2 polypeptide.
XX
KW Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
KM gene therapy; protein therapy.
XX
OS Homo sapiens.
XX
PN WO200000605-A1.
XX
PD 06-JAN-2000.
XX
PE 29-JUN-1999; 99WO-US14622.
XX
PR 29-JUN-1998; 98US-0091044.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Hong AKC, Tavtliglan SV, Teng DHF;
XX
DR WPI: 2000-170914/15.
XX
DR N-PSDB; AA287786.
XX
PT Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of
PT human cancer -
```

```
XX PS Claim 55; Page 77-79; 89pp; English.
XX
CC The invention provides a new tumour suppressor gene, designated TMPRSS2.
CC The TMPRSS2 polynucleotides and polypeptides can be used in methods for
CC diagnosing and prognosing predisposition to cancer in humans. The
CC polypeptides may also be used in assays to screen for compounds with
CC anti-cancer or therapeutic properties. The polypeptides are also useful
CC for rational drug design. The TMPRSS2 polynucleotides and polypeptides
CC may be used for gene therapy and protein therapy. The present sequence
CC represents the TMPRSS2 polypeptide.
XX
SQ Sequence 492 AA;

Query Match          100.0%; Score 1319; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.4e-117;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPWOVSLHVQNVHVCGSITTPEMWITAAHCVEKPLNPMHTAFAGI 60
DB 255 RIVGESALPGAMPWOVSLHVQNVHVCGSITTPEMWITAAHCVEKPLNPMHTAFAGI 314
OY 61 LRQSFMYGAGYQVEKYISHPNYDSKTRKNDIALMKLOKPLTFNDLYKPVCLPMPGMQLQ 120
DB 315 LRQSFMYGAGYQVEKYISHPNYDSKTRKNDIALMKLOKPLTFNDLYKPVCLPMPGMQLQ 374
OY 121 PEOLCWISMGATEEKGTEKTSVELNAKVLLETQRCNSRYVYDNLITPAMICAGFLQGNV 180
DB 375 PEOLCWISMGATEEKGTEKTSVELNAKVLLETQRCNSRYVYDNLITPAMICAGFLQGNV 434
OY 181 DSCGDSGGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGVNMFDMITRQMRADG 238
DB 435 DSCGDSGGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGVNMFDMITRQMRADG 492

RESULT 3
AAV44406
ID AAV44406 standard; Protein: 492 AA.
XX
AC AAV44406;
XX
DT 22-MAR-2000 (first entry)
XX
DE Human 20P1F12-GTC2 protein.
XX
KW 20P1F12; TMPRSS2; androgen; serine protease; 20P1F12-GTC1; cancer;
KM transmembrane protein; colon; prostate; prostate tumour.
XX
OS Homo sapiens.
XX
PN WO9962942-A2.
XX
PD 09-DEC-1999.
XX
PE 01-JUN-1999; 99WO-US12253.
XX
PR 01-JUN-1998; 98US-0087598.
XX
PR 29-JUN-1998; 98US-0091474.
XX
PR 14-APR-1999; 99US-0129521.
XX
PA (UROG-) UROGENESTS INC.
XX
PA (AFAR/) AFAR D E.
XX
PA (HUBE/) HUBERT R S.
XX
PA (LEON/) LEONG K.
XX
PA (RAIT/) RAITANO A B.
XX
PA (SAFF/) SAFFRAN D C.
XX
PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;
XX
DR WPI: 2000-116363/10.
XX
DR N-PSDB; AA229636.
XX
```

PF Novel cell surface antigen useful to treat colon and prostate cancer -
 XX
 PS Claim 1, Fig 1, 58pp; English.
 XX

CC The present sequence is the 20P1F12 protein (also known as the TMPRSS2
 CC protein) which is a prostate-specific, androgen-regulated, cell surface
 CC serine protease. It is a glycosylated type II transmembrane protein with
 CC an extracellular C-terminal serine protease domain, a scavenger receptor
 CC cysteine-rich domain, an LDL receptor class A domain and a predicted
 CC transmembrane domain. Host cells can be transformed to produce this
 CC protein, using vector containing 20P1F12/TMPRSS2 gene (also designated
 CC 20P1F12-Gfcl, as deposited with ATCC accession number 207097).
 CC Anti-20P1F12/TMPRSS2 antibodies may be used as therapeutic agent for
 CC prostate and colon cancers, to image prostate cancer cells and
 CC prostate tumours, to identify ligands and cellular constituents that
 CC bind to a 20P1F12/TMPRSS2 gene product and for use as cancer vaccines.
 CC

XX Sequence 492 AA:

Query Match 100.0%; Score 1319; DB 21; Length 492;
 Best Local Similarity 100.0%; Pred. No. 5.4e-117;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITPEWITVAHCVEKPLNNPMTAFAGI 60
 DB 255 rivgesalpgampqvslhvqnvhvcgsiltpewivaahcveklpnnpmtafagi 314
 QY 61 LRQSFMEYGAGYQVEKVISHPNYSKTKNDIALMKLQKPLTFNDLVKPCLPNGMMLQ 120
 DB 315 lrqsfmefyagayqvekvishpnysktnkdialmklqkpltfndlvkpcldpnmmlq 374
 QY 121 PRLQWISGMGATTEEGKSEVILNAKVLITFORNSRYVDNLITPMICAGFLQGNV 180
 DB 375 prqlcwisgwgatteeekgtsevinakvllitforcnstryvndltpmicagflqgnv 434
 QY 181 DSCQGDGSGPLVTSKNNIMWLIGDTSMSGSCAKAYRPYGVNVMFTDWTYROMRADG 238
 DB 435 dscqgdsgsplvtstknimwlligdtswsgscakayrpygvnmvftdwtlyqmrtdg 492

RESULT 4
 ID AAY81492 standard; Protein; 283 AA.
 XX AAY81492;
 AC
 XX 18-JUL-2000 (first entry)
 DT
 XX Human prostate-associated protease (HUPAP).
 DE
 XX Human prostate-associated protease; HUPAP; kallikrein; serine protease;
 KM gastrointestinal disorder; cancer; prostate disorder.
 KW
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT Modified-site 4 /note= "N-glycosylated"
 FT Misc-difference 235 /label= unknown
 FT /note= "Encoded by CNT"
 FT
 XX US6043033-A.
 PN
 XX 28-MAR-2000.
 PD
 XX 27-FEB-1997; 97US-0807151.
 PF
 XX 27-FEB-1997; 97US-0807151.
 PR
 XX (INCY-) INCYTE PHARM INC.
 PA
 XX

PI Bandman O, Lal P;
 XX
 XX WPI: 2000-28523/24.
 DR N-PSDB: AAA12975.
 XX

PT Polynucleotide encoding human prostate-associated protease useful for
 PT diagnosing and treating cancers, prostate disorders and
 PT gastrointestinal disorders -
 PS Claim 1; Fig 1A-B; 27pp; English.
 PS

CC This sequence represents human prostate-associated protease (HUPAP). cDNA
 CC encoding HUPAP was initially identified in a spinal cord cDNA library,
 CC the cDNA encoding this sequence representing a consensus of overlapping
 CC and/or extended nucleic acid sequences from spinal cord, prostate tumour
 CC and/or cDNA libraries. HUPAP is a serine protease with structural and
 CC functional homology with bovine enterokinase, human pancreatic kallikrein
 CC and African rat renal kallikrein, sharing 38% homology with bovine
 CC enterokinase. In addition, HUPAP is related to prostate-specific antigen
 CC (PSA), a kallikrein which is a highly sensitive marker for prostate
 CC cancer. HUPAP and nucleic acids encoding it are useful for the diagnosis,
 CC prevention and treatment of gastrointestinal disorders such as ulcerative
 CC colitis, pancreatitis, cancers, and prostatic disorders. HUPAP agonists
 CC may be used to treat gastrointestinal disorders, and HUPAP antagonists
 CC and inhibitors may be used to suppress excessive cell proliferation,
 CC which is of use in cancer therapy. HUPAP nucleic acids are also useful
 CC for generating hybridisation probes which may be used for mapping
 CC naturally occurring genomic sequences.
 CC

XX Sequence 283 AA:

Query Match 99.3%; Score 1310; DB 21; Length 283;
 Best Local Similarity 99.6%; Pred. No. 2e-116;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITPEWITVAHCVEKPLNNPMTAFAGI 60
 DB 46 rivgesalpgampqvslhvqnvhvcgsiltpewivaahcveklpnnpmtafagi 105
 QY 61 LRQSFMEYGAGYQVEKVISHPNYSKTKNDIALMKLQKPLTFNDLVKPCLPNGMMLQ 120
 DB 106 lrqsfmefyagayqvekvishpnysktnkdialmklqkpltfndlvkpcldpnmmlq 165
 QY 121 PRLQWISGMGATTEEGKSEVILNAKVLITFORNSRYVDNLITPMICAGFLQGNV 180
 DB 166 prqlcwisgwgatteeekgtsevinakvllitforcnstryvndltpmicagflqgnv 225
 QY 181 DSCQGDGSGPLVTSKNNIMWLIGDTSMSGSCAKAYRPYGVNVMFTDWTYROMRADG 238
 DB 226 dscqgdsgsplvtstknimwlligdtswsgscakayrpygvnmvftdwtlyqmrtdg 283

RESULT 5
 ID AAB36901 standard; Protein; 492 AA.
 XX AAB36901;
 AC
 XX 26-FEB-2001 (first entry)
 DT
 XX Human TMPRSS2 protein.
 DE
 XX Prostate specific androgen regulated protein; ARSDR1; TMPRSS2;
 KM PART-1; neoplastic.
 KW
 XX Homo sapiens.
 OS
 XX WO200065067-A2.
 PN
 XX 02-NOV-2000.
 PD
 XX 21-APR-2000; 2000WO-US10920.
 PF

XX 23-APR-1999; 99US-0130778.
PR 30-AUG-1999; 99US-0151585.
PR 30-DEC-1999; 99US-0174003.
PR 24-JAN-2000; 2000US-0177751.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Nelson PS, Hood L, Lin B;
XX
DR WPI: 2000-679676/66.
DR N-PSDB: AAC83325.
XX
PT Polynucleotide encoding prostate specific androgen regulated
PT polypeptides and inhibitor of the peptides useful for treating or
PT reducing the progression of prostate neoplastic condition in an
PT individual -
XX
PS Claim 63; Page 86-88; 121pp; English.
XX
PS The present invention relates to prostate specific androgen regulated
CC proteins. The invention may be used to determine an expression level
CC of the prostate-specific proteins ARSDR, TMPRSS2, or PART-1 in a
CC fluid sample or prostate cell sample from an individual. It may also
CC be used for diagnosing and predicting the susceptibility of a
CC prostate neoplastic condition in an individual. Inhibitors of the
CC proteins are useful for treating or preventing the progression of a
CC prostate neoplastic condition.
XX
SQ Sequence 492 AA;

Query Match 98.8%; Score 1303; DB 21; Length 492;
Best Local Similarity 98.3%; Pred. No. 1.8e-115;
Matches 234; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 RIVGESALPGAMPQVSLHVNHVCGSITTPWITYTAHCVEKPLNPNHMTAFAGI 60
DB 255 RIVGESALPGAMPQVSLHVNHVCGSITTPWITYTAHCVEKPLNPNHMTAFAGI 314
QY 61 LROSFMFGAGYQVQKVSHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMMLQ 120
DB 315 LRSFMTFYSAGYQVQKVSHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMMLQ 374
QY 121 PEOLCWMISGNGATEEKGKTSYVLAHAKVLLIETQRCNSRYVDNLITPAMICAGFLQGNV 180
DB 375 PEQLCWMISGNGATEEKGKTSYVLAHAKVLLIETQRCNSRYVDNLITPAMICAGFLQGNV 434
QY 181 DSCGDSGGLPVTSTKNNIWMILGDTSMGSCAKATRPVYGVNVFTDWTIRQMRADG 238
DB 435 DSCGDSGGLPVTSTKNNIWMILGDTSMGSCAKATRPVYGVNVFTDWTIRQMRADG 492

RESULT 6
AA57280
ID AA57280 standard; Protein: 492 AA.
XX
AC AA57280;
XX
DT 06-JUN-2000 (first entry)
XX
DE Ovr115 homolog protein.
XX
KW CS6; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;
KW endometrial; uterine; lung; cytotoxic.
XX
OS Homo sapiens.
XX
PN MO200012758-A1.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99MO-US19655.

XX 02-SEP-1998; 98US-0098880.
XX
XX (DIAD-) DIADEXUS LLC.
XX
PI Salceda S, Sun Y, Reclon H, Caferkey R;
XX
DR WPI: 2000-256657/22.
DR N-PSDB: AA290478.
XX
PT Diagnosing, staging, monitoring, imaging and treating cancer especially
PT gynecological cancers e.g. breast, ovarian cancer and lung cancer,
PT involves measuring cancer specific gene levels in cells and body fluids
PT
XX
PS Disclosure; Page 52-54; 58pp; English.
XX
XX The invention relates to detecting, diagnosing metastasis and staging
CC cancer by measuring levels of cancer specific genes (CSG) in cells,
CC tissues or body fluids. Their remission and progression, decreases and
CC increases in CSG levels, is also monitored, by periodic sample analysis.
CC The methods are useful for detecting cancers, especially gynecologic
CC cancers which include ovarian, breast, endometrial and uterine cancer
CC and lung cancer. Antibodies against the CS6s labeled with paramagnetic
CC ions or a radioisotope are useful for imaging cancer and when conjugated
CC with a cytotoxic agent are useful for treating cancer. The present
CC sequence represents a Ovr115 homolog protein, that can be used
CC for the detection of the various cancers.
XX
SQ Sequence 492 AA;

Query Match 98.8%; Score 1303; DB 21; Length 492;
Best Local Similarity 98.3%; Pred. No. 1.8e-115;
Matches 234; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 RIVGESALPGAMPQVSLHVNHVCGSITTPWITYTAHCVEKPLNPNHMTAFAGI 60
DB 255 RIVGESALPGAMPQVSLHVNHVCGSITTPWITYTAHCVEKPLNPNHMTAFAGI 314
QY 61 LROSFMFGAGYQVQKVSHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMMLQ 120
DB 315 LRSFMTFYSAGYQVQKVSHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPDPGMMLQ 374
QY 121 PEOLCWMISGNGATEEKGKTSYVLAHAKVLLIETQRCNSRYVDNLITPAMICAGFLQGNV 180
DB 375 PEQLCWMISGNGATEEKGKTSYVLAHAKVLLIETQRCNSRYVDNLITPAMICAGFLQGNV 434
QY 181 DSCGDSGGLPVTSTKNNIWMILGDTSMGSCAKATRPVYGVNVFTDWTIRQMRADG 238
DB 435 DSCGDSGGLPVTSTKNNIWMILGDTSMGSCAKATRPVYGVNVFTDWTIRQMRADG 492

RESULT 7
AA41694
ID AA41694 standard; Protein: 452 AA.
XX
AC AA41694;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO382 protein sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
OS Homo sapiens.
XX
PN MO9946281-A2.
XX
PD 16-SEP-1999.
XX

DT 11-JAN-2001 (first entry)
 XX Tumour associated differentially-expressed gene 12 protein sequence.
 DE
 XX Transmembrane serine protease; TADG-12; chromosome 17; vaccination;
 KW tumour associated differentially-expressed gene 12; cytosolic; human;
 KW malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
 XX
 OS Homo sapiens.
 XX
 PN MO200052044-A1.
 XX
 PD 08-SEP-2000.
 XX
 PF 02-MAR-2000; 2000WO-US05612.
 XX
 PR 03-MAR-1999; 99US-0261416.
 XX
 PA (UYAR-) UNIV ARKANSAS.
 PI O'Brien TJ, Underwood LJ;
 XX
 DR WPI: 2000-533263/48.
 DR N-PSDB: AAA93842.
 XX
 PT DNA fragment encoding tumor associated differentially-expressed gene 12
 PT protein used for diagnosing and treating malignant hyperplasia and
 PT cancers including ovarian cancer -
 XX
 PS Claim 3; Figure 4; 118bp; English.
 XX
 CC This invention relates to a novel transmembrane serine protease called
 CC tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is
 CC located on chromosome 17. Sequences AAA93842-A93845 and AAB32246-B32249
 CC represent human TADG-12 cDNA and their corresponding protein sequences.
 CC A splice variant of TADG-12 (TADG-12V) leads to a truncated protein
 CC product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12
 CC exhibits cytosolic activity, and can be used in vaccines and in gene
 CC therapy. TADG-12 nucleotide and protein sequences are used in the
 CC diagnosis of malignant hyperplasia and cancers of the ovary, breast,
 CC lung, colon, prostate and other cancers where TADG-12 is overexpressed.
 CC TADG-12 is particularly used as tumour marker for early disease
 CC diagnosis. TADG12 proteins or fragments can be used to vaccinate an
 CC individual with cancer, suspected of having a cancer or at risk of
 CC getting cancer. Sequences AAA93846-A93853 represent PCR primers used for
 CC amplifying the TADG-12 cDNA sequence, and in the quantitative analysis of
 CC TADG-12 mRNA. AAB32250 represents a peptide fragment of TADG-12, used to
 CC create anti-TADG-12 antibodies. Sequences AAB32251-B32369 represent
 CC TADG-12 peptides which target HLA, and may be used in a vaccine or for
 CC immune stimulation.
 CC
 XX
 SQ Sequence 454 AA;

Query Match 49.7%; Score 655; DB 21; Length 454;
 Best Local Similarity 52.9%; Pred. No. 4.5e-54;
 Matches 126; Conservative 30; Mismatches 78; Indels 4; Gaps 3;

QY 1 RIVGESALPGAMPQVSLHYQNVHVCSSITPEWITAHACVKEKPLNNPMHTAFAGI 60
 DB 217 RIVGSMSSLLSQPWQASLQFGYHLCGGSVILPLWITAAHCV-YDLIYKSVTLQYGI 275
 QY 61 LROSFMEVGA-GYQVEKVISHPNYSKRNNDIALMLKQLKPLTFNDLKYVCLPNPGMML 119
 DB 276 V--81ldnppshlveklyvksykprkrgndialmklagpltfnemlqpcvclpnsenf 333
 QY 120 QPEOLCHISGKATFEKGTSEVMAAKVLLIETQRCSKRVYDNLITPAITCAFLQGN 179
 DB 334 pdgkvvcwsgatcdgdaepvlnhaavpliankicnhrdvvgyslspnalcagyltg 393
 QY 180 VDSGCGSGGLPVTSKNNIMWLIGDTSWGSCAKAYRPGVGNVWPFDMVIRQMRAD 237
 DB 394 vdsqgdsagpplvcqerllwklvgatstlgcaevnkpvytlrvtsfldwlnheqmer 451

RESULT 9
 AAB44250
 ID AAB44250 standard; Protein: 453 AA.
 XX
 AC AAB44250;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human PRO382 (UNQ323) protein sequence SEQ ID NO:69.
 XX
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytosolic;
 KW expressed sequence tag; detection; cancer.
 XX
 OS Homo sapiens.
 XX
 PN MO200053756-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 18-FEB-2000; 2000WO-US04341.
 XX
 PR 08-MAR-1999; 99WO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL,
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME,
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ,
 PI Kljavin LJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA,
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 XX
 DR WPI: 2000-611443/58.
 DR N-PSDB: AAC78475.
 XX
 PT Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities -
 XX
 PS Claim 12; Fig 28; 636bp; English.

AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 sequence tag) sequences which encode secreted or transmembrane PRO
 polypeptides. The PRO polynucleotides and polypeptides have cytosolic
 activity. The polynucleotides and polypeptides can be used for detecting
 the presence of PRO polypeptides in samples, for linking bioactive
 molecules to cells and for modulating biological activities of cells.
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.
 CC The polypeptide pairs provide specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.
 XX
 SQ Sequence 453 AA;

RESULT	10
AA043572	
ID	AA043572 standard; Protein; 248 AA.
XX	
XX	AA043572;
AC	
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human cancer associated protein sequence SEQ ID NO:1017.

KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnereary; immunomodulator;
 KW antidiabetic; antiaesthatic; antirheumatic; antiarthritis; antiviral;
 KW antineoplastic; antitumor; antitumor; antitumor; antitumor;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
 KW vasotropic; antiproliferative; antiangiogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening;
 XX
 OS Homo sapiens.
 XX
 PN WO200055350-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05882.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;.
 XX
 DR WPI: 2000-587533/55.
 DR N-PSDB: AAC77781.
 XX
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 XX useful for treating or diagnosing e.g. cancer -
 XX
 PS Claim 11; Page 1598-1599; 2352pp; English.

CC AAC767607.0 AAC78448 encode the human cancer associated proteins given
CC in AAB33368 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytosstatic; proliferative; vulnerrary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatologic; neuroprotective; cardiac; thrombolytic; coagulant;
CC neurotropic; vasotropic; antiprolastic and antiangiogenic. CC

Query Match	48.9%	Score 645.5	DB 21	Length 248
Best Local Similarity	52.7%	Pred. No. 1.7e-53		
Matches 126; Conservative	30;	Mismatches 78;	Indels 5;	Gaps 4;

```

QY 1 RIVGESALIPGAMPVOVSLHNVONVCGSSITTPRMITAALCHYEKRLPNMHHWAFAGI 60
Db 10 rlvvgmslslsqwpqasqlfggylrcgsvsltrplwiltlaacv-ydylrlpswtlgyvl 68
QY 61 LRQSMFVYGA-VGYEYKVLISHPNYDSKTEPNNDIALMLKLOKPLTFENDLVKPYCLPPMGML 119
Db 69 v--slldnpapshlyeklyvshkypkrlgndialmlktagplrtfremtqgvclpseenf 126
QY 120 OPEOLCWTISMGATEE-KGKTSEVLNAAKVLLETRQCRNSRYVYDNLITTPAMICAGFLQG 178
Db 127 pdkgkcwtsywgatedsgdgdaasvlnhaavplnsnklichrdvygslslspsmlcsqyltg 186
QY 179 NVDSCQGBSGGGLVYSKNNIMWLIGDTISWSSCAALAPRGVYGNNVWTTDWIYRQMRAD 237
Db 187 gvdscqgsgsgglvcqerrlwlklyvatsfglsgvavnpypvylrvtlsldwlshegmed 245

```

RESULT	11
AA72093	
ID	AA72093 standard; Protein: 327 AA.

AC	AAV72093,
XX	
DT	28-MAR-2001 (first entry)
XX	
DE	Human serine protease #4 encoded by clone HCHAK72.

KM Human; stine protease; osteopathic; immunosuppressive; antiallergic;
 KM antiinflammatory; cytostatic; cardiac; neuroprotective; nootropic;
 KM neuroleptic; vitruvery; ophthalmological; antibacterial; antiviral;
 KM antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
 KM treatment; bone formation disorder; osteoporosis; arthritis; cancer;
 KM connective tissue disorder; autoimmune disorder; wound healing; asthma;
 KM systemic lupus erythematosus; male reproductive system disorder;
 KM testicular cancer; digestion and food absorption disorder; arrhythmia;
 KM Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
 KM behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
 KM cardiovascular disorder; ocular disorder; drug screening.

OS Homo sapiens.
XX
XX W0200068247-A2.
PN
XX
XX
PD 16-NOV-2000.
XX
XX
PF 05-MAY-2000; 2000WO-US12207.
XX
XX
PR 07-MAY-1999; 99US-0133239.
PR 20-MAY-1999; 99US-0135163.
PR 03-AUG-1999; 99US-0147005.
PR 09-SEP-1999; 99US-0152935.
PR 01-NOV-1999; 99US-0162979.

XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Shi Y, Young PE, Ni J;
PI WPI: 2000-679799/66.
XX N-PSDB: AAD02323.
DR
XX
XX New nucleic acid molecules encoding human serine protease polypeptides,
PT useful for diagnosis, prevention and/or treatment of disorders e.g.
PT osteoporosis, lupus erythematosus and Alzheimer's -
PS
XX
XX Claim 12: Page 280-281; 289pp: English.
XX
XX The present sequence is human serine protease #4 from clone
CC HCHAK72 (ATCC Deposit No: PTA27).
CC The invention relates to human serine proteases and their cDNA clones.
CC It is used in methods for the diagnosis, prevention and treatment of
CC various disorders related to serine protease such as bone formation
CC disorders (osteoporosis), connective tissue disorders (arthritis),
CC autoimmune disorders (systemic lupus erythematosus), wound healing, male
CC reproductive system disorders (testicular cancer), digestion and food
CC absorption disorders (Crohn's disease), neurodegenerative diseases
CC (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
CC proliferative and cancerous conditions (acute myelogenous leukemia),
CC allergic reactions (asthma), cardiovascular disorders (arrhythmia),
CC ocular disorders (glaucoma) and infectious diseases caused by bacteria,
CC viruses, fungi or parasites. It is also useful for screening therapeutic
CC compounds. Serine proteases are used as immunological probes or
CC polymorphic markers for the identification of chromosomes, cells and
CC tissues in biological samples, identification of male contraceptive
CC agents, delivery of compositions to targeted cells expressing a
CC receptor for serine protease, hybridisation probes and molecular weight
CC markers. Serine protease nucleic acids are also useful in gene therapy.
CC Note: The present sequence shown in page 280-281 of sequence listing has
CC been assigned SEQ ID NO: 14. But the sequence, human serine protease
CC epidermal growth factor (EGF)-like domain (AAV2114) shown in
CC page 12 is also referred as SEQ ID NO: 14.
XX
XX
XX Sequence 327 AA;
SQ
Query Match 48.9%; Score 645.5; DB 21; Length 327;
Best Local Similarity 52.7%; Pred. No. 2.4e-53;
Matches 126; Conservative 30; Mismatches 78; Indels 5; Gaps 4;
QY 1 RIVGESALPGAMPQVSLHVNQVHVCSSITTPMIVTAACVCEKPLNPNMHTAFAGI 60
DB 89 RIVGSMMLISGVPWQGAALGFQGYHLGCGSVLTPLWILTAACVYDLYLPKSWELQYGI 147
QY 61 LRQSFMEYGA-GYQVEKVISHPNYSKTKNNDIALMKLQKPLTFNDLVKPYCLPMPGKML 119
DB 148 V--allndpaphllyveklivhskypkrlgndlaiaimklagpltfnemlqpvclpseeenf 205
QY 120 QPEQLCHISGNGATEE-KGKTSSEVLNAKVLLEIQRNRSRYVNDLTPAMICAGFLQG 178
DB 206 pdgkvcwsgvsgatadgsdpsvlnhaavplsnklnchtdvvgllspnmlcsqyltg 265
QY 179 NVDSQGDGSGPLVYTSKNNIMWLLIGDTSMSGSCAKARPGYGVGNVMTDWTIMRMD 237
DB 266 gydsqsgsgplvcqdrllwklvgatstfgcgaevnkpvytlrtstfldwhegmrd 324
RESULT 12
AAB08912
ID AAB08912 standard; Protein: 414 AA.
XX
XX AAB08912;
AC
XX
XX 30-AUG-2000 (first entry)
XX
XX Human secreted protein sequence encoded by gene 22 SEQ ID NO:69.
DE

KW Human; secreted protein; cytosolic; anti-proliferative; vulnery;
KM immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
KM hyperproliferative disorder; infectious disease; tissue regeneration;
KM screening; food additive; preservative; wound healing;
KM hyper-vascular disease.
OS
XX Homo sapiens.
PN WO200017222-A1.
XX
XX 30-MAR-2000.
PD
XX
XX 22-SEP-1999; 99WO-US22012.
XX
XX 23-SEP-1998; 98US-0101546.
PR 02-OCT-1998; 98US-0102895.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
PI Komatsoulis G, Endress GA, Soppet DR;
XX N-PSDB: AAA39073.
DR
XX WPI: 2000-283538/24.
PS
XX
XX Claim 11: Page 367-368; 416pp: English.
XX
XX The polynucleotide sequences given in AAA39052 to AAA39088 encode the
XX human secreted proteins given in AAB08891 to AAB08984. The human secreted
XX proteins can have activities based on the tissues and cells they are
XX expressed in. Examples of the activities are: cytostatic;
XX anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The
XX secreted proteins and their related polynucleotide sequences are useful
XX for diagnostic and therapeutic methods useful for diagnosing and treating
XX disorders related to the secreted proteins. The proteins, and
XX polynucleotide sequences may be useful for treating disorders of the
XX immune system, hyperproliferative disorders, infectious disease,
XX regeneration of tissues, for chemotaxis and for screening molecules that
XX bind to the proteins. The proteins or polynucleotide sequences may be
XX used as food additives or preservatives. To increase or decrease storage
XX capabilities, fat content, lipid, protein, carbohydrate, vitamins,
XX minerals, co-factors or other nutritional components. Agents or
XX antagonists of the proteins may be used to prevent scar tissue growth
XX during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
XX and AAB08890 are sequences used in the exemplification of the present
XX invention.
SQ
XX
XX Sequence 414 AA;
Query Match 42.8%; Score 564; DB 21; Length 414;
Best Local Similarity 45.3%; Pred. No. 1.7e-45;
Matches 106; Conservative 31; Mismatches 94; Indels 2; Gaps 2;
QY 1 RIVGESALPGAMPQVSLHVNQVHVCSSITTPMIVTAACVCEK-PLNPNMHTAFAG 59
DB 173 RIVGSGVAPGRVPQGAVALGFRTCGSVLAPRVVLTAAHCHMSFIALALSSRYRHAG 232
QY 60 ILRQSFMEYGAQYQVEKVISHPNYSKTKNNDIALMKLQKPLTFNDLVKPYCLPMPGKML 119
DB 233 lvshaavrpghalverlliplysaqnhdydvalrlrldtalsdvtgavclpakeghf 292
QY 120 QPEQLCHISGNGATEEKGK-TSEVLNAKVLLEIQRNRSRYVNDLTPAMICAGFLQG 178
DB 293 pkgsrvcwsgvghthpshlycsdmldtvpdlfstqclnscscvsgaltprmlcaqyltg 352
QY 179 NVDSQGDGSGPLVYTSKNNIMWLLIGDTSMSGSCAKARPGYGVGNVMTDWTIMY 231
DB 353 radacqsgsgplvcpgdgtlwrivgvswwrgcdephnpvyakvaefldwln 405

RESULT 13
AAB08950
ID AAB08950 standard; Protein; 480 AA.
XX
AC AAB08950;
XX
DT 30-AUG-2000 (first entry)
XX
DE Human secreted protein sequence encoded by gene 22 SEQ ID NO:107.
XX
KW Human; secreted protein; cytosolic; anti-proliferative; vulnery;
KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
KW hyperproliferative disorder; infectious disease; tissue regeneration;
KW screening; food additive; preservative; wound healing;
KW hyper-vascular disease.
XX
XX Homo sapiens.
OS
PN MO200017222-A1.
XX
PD 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-US22012.
XX
PR 23-SEP-1998; 98US-0101546.
PR 02-OCT-1998; 98US-0102895.
XX
PA (HDMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
PI Komatsoulis G, Endress GA, Soppet DR;
PI WPI: 2000-283538/24.
XX
DR Human secreted proteins and coding sequences useful in diagnostic and
PT therapeutic methods for disorders such as immune system or
PT proliferative disorders, related to the proteins
XX
PS Disclosure: Page 65; 416pp; English.
XX
CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC human secreted proteins given in AAB08981 to AAB08984. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC anti-proliferative; immunosuppressive; antibacterial; and vulnery. The
CC secreted proteins and their related polynucleotide sequences are useful
CC for diagnostic and therapeutic methods useful for diagnosing and treating
CC disorders related to the secreted proteins. The proteins, and
CC polynucleotide sequences may be useful for treating disorders of the
CC immune system, hyperproliferative disorders, infectious disease,
CC regeneration of tissues, for chemotaxis and for screening molecules that
CC bind to the proteins. The proteins or polynucleotide sequences may be
CC used as food additives or preservatives, to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, co-factors or other nutritional components. Agonists or
CC antagonists of the proteins may be used to prevent scar tissue growth
CC during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
CC and AAB08980 are sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 480 AA:
XX
Query Match 42.8%; Score 564; DB 21; Length 480;
Best Local Similarity 45.5%; Pred. No. 2e-45;
Matches 106; Conservative 31; Mismatches 94; Indels 2; Gaps 2;
QY 1 RIVGGSALPGAMPQVSLHVNQVHCGSITPEWIVTAHAKVER-PLNPNPHMTAFAG 59
DB 240 RIVGGSVAPGTPWQVAVSLGFRHCGSVLAPRWVLAHCHMSFRLARISWTVHAG 299

QY 60 ILRQSFMYGAGYQVEKVISHPNYSKTRKNNDIALKLOKPLTFNDLVPVCLPNPMM 119
DB 300 lvsasvrbphgavlerilphlypsagmndyavallrlqtafnfstvavclpaqehf 359
QY 120 QPEQLCWISGWCATPEERK-TSEVLNAKAVLLIETORCSRYYVNLITPAMICAFLOG 178
DB 360 pkgrcwsvsgwghthpshtyscdmlqdtvplfstqlcnsscvysgalprmlcagylg 419
QY 179 NVDSGCGDSCGLVYTSKNNIWMILIGDTSMGSCGAKAVRGYGVNVPFDMY 231
DB 420 radacqdsqgplvcpgdltwrlivgvswgqcaepnhpyakvaeFlawh 472
RESULT 14
AAY72108
ID AAY72108 standard; Protein; 296 AA.
XX
AC AAY72108;
XX
DT 28-MAR-2001 (first entry)
XX
DE Human serine protease #3 fragment from HATEB38 cDNA clone.
XX
KW Human; serine protease; osteopathic; immunosuppressive; anti-allergic;
KW anti-inflammatory; cytostatic; cardiant; neuroprotective; nocrotic;
KW neuroleptic; vulnery; ophthalmological; antibacterial; antiviral;
KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
KW systemic lupus erythematosus; male reproductive system disorder;
KW testicular cancer; digestion and food absorption disorder; arrhythmia;
KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
KW behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
KW cardiovascular disorder; ocular disorder; drug screening.
XX
XX Homo sapiens.
OS
XX
PN MO200068247-A2.
XX
PD 16-NOV-2000.
XX
PF 05-MAY-2000; 2000WO-US12207.
XX
PR 07-MAY-1999; 99US-0133239.
PR 20-MAY-1999; 99US-0135163.
PR 03-AUG-1998; 99US-0147005.
PR 09-SEP-1999; 99US-0152935.
PR 01-NOV-1999; 99US-0162979.
XX
PA (HDMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Shi Y, Young PE, Ni J;
PI WPI: 2000-679799/66.
PI N-PSDB: AAD02322.
XX
DR New nucleic acid molecules encoding human serine protease polypeptides,
PT useful for diagnosis, prevention and/or treatment of disorders e.g.
PT osteoporosis, lupus erythematosus and Alzheimer's
XX
PS Disclosure: Page 286-287; 289pp; English.
XX
CC The present sequence is human serine protease #3 fragment from clone
CC HATEB38 (ATCC Deposit No. PTA27).
CC The invention relates to human serine proteases and their cDNA clones.
CC It is used in methods for the diagnosis, prevention and treatment of
CC various disorders related to serine protease such as bone formation
CC disorders (osteoporosis), connective tissue disorders (arthritis),
CC autoimmune disorders (systemic lupus erythematosus), wound healing, male
CC reproductive system disorders (testicular cancer), digestion and food
CC absorption disorders (Crohn's disease), neurodegenerative diseases
CC (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
CC proliferative and cancerous conditions (acute myelogenous leukaemia),

THIS PAGE BLANK (USPTO)

Gencore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:42 : Search time 25.88 Seconds

(without alignments)
700.523 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492

Perfect score: 1319

Sequence: 1 RYGGESALPGAMPQVSLH.....VYGNVWFDTQWYRQNRADG 238

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: PIR68:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	558.5	42.3	638	1	KOHUP plasma kallikrein
2	543	41.2	638	1	KORTPL plasma kallikrein
3	530	40.2	638	1	KOMSPPL plasma kallikrein
4	529	40.1	625	1	KFHUI coagulation factor
5	528.5	40.1	417	1	S00845 hepsin (EC 3.4.21.
6	523.5	39.7	1035	1	A43090 enteropeptidase (E
7	513.5	38.9	1019	1	A55318 enteropeptidase (E
8	502.5	38.1	812	1	PLMS plasma (EC 3.4.21
9	499.5	37.9	416	1	S33777 hepsin (EC 3.4.21.
10	499	37.8	1524	2	T30337 polypeptide - Affi
11	492	37.3	421	1	S11674 acrosin (EC 3.4.21
12	488.5	37.0	1034	1	A53663 enteropeptidase (E
13	481.5	36.5	415	1	A34170 acrosin (EC 3.4.21
14	476.5	36.1	436	2	UX0172 acrosin (EC 3.4.21
15	466.5	35.4	431	2	S47538 acrosin (EC 3.4.21
16	464.5	35.2	421	2	S29599 acrosin (EC 3.4.21
17	464	35.2	418	2	A37344 acrosin (EC 3.4.21
18	464	35.2	790	1	PLPG plasma (EC 3.4.21
19	460.5	34.9	437	2	S18407 acrosin (EC 3.4.21
20	458	34.7	810	2	B30848 plasmin (EC 3.4.21
21	457.5	34.7	810	2	I46260 plasmin (EC 3.4.21
22	454	34.4	460	2	B61545 plasmin (EC 3.4.21
23	451.5	34.2	812	1	PLBO plasmin (EC 3.4.21
24	450.5	34.2	267	2	S40006 plasmin (EC 3.4.21
25	450	34.1	270	2	S56160 mast cell tryptase
26	449	34.0	275	2	S40005 trypsin (EC 3.4.21
27	449	34.0	343	1	A57014 prostaasin (EC 3.4.
28	447	33.9	276	2	A38654 mast cell proteina
29	447	33.9	277	2	S35340 trypsin (EC 3.4.21

30	443	33.6	455	2	A61545 plasmin (EC 3.4.21
31	442.5	33.5	266	2	S54146 trypsin (EC 3.4.21
32	442.5	33.5	4548	1	S00657 apoprotein(a) (EC
33	442	33.5	810	1	PLMU plasmin (EC 3.4.21
34	441	33.4	1113	2	JE0315 low-density lipopr
35	440	33.4	274	2	JC4171 trypsin (EC 3.4.2
36	439	33.3	273	2	A47246 trypsin (EC 3.4.2
37	438	33.2	274	2	S35339 trypsin (EC 3.4.21
38	437	33.1	275	2	S40007 trypsin (EC 3.4.21
39	433.5	32.9	420	2	A55283 acrosin (EC 3.4.21
40	432	32.8	242	2	S49489 trypsin (EC 3.4.21
41	432	32.8	275	2	C35863 trypsin (EC 3.4.2
42	430	32.6	244	2	S72219 chymotrypsin B - A
43	430	32.6	275	2	A35863 trypsin (EC 3.4.2
44	429	32.5	263	1	KYRTB chymotrypsin (EC 3
45	429	32.5	275	2	B35863 trypsin (EC 3.4.2

ALIGNMENTS

RESULT 1

KOHUP
plasma kallikrein (EC 3.4.21.34) precursor - human
M:Alternate names: kininogenin; plasma prekallikrein
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1986 #sequence, revision 13-Aug-1986 #text, change 18-Jun-1999
C:Accession: A00921; A57939
R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains fou
A:Reference number: A00921; MUID:86243359
A:Accession: A00921

A:Molecule type: mRNA
A:Residues: 1-638 <CHD>
A:Cross-references: GB:M13143; NID:9190262; PIDN:AAA60153.1; PID:9190263
A:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence
A:Reference number: A37939; MUID:91152016
A:Accession: A37939

A:Molecule type: protein
A:Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;1
;260-283,'X',285;287-291,'X',293-295;314-317,'X',319-320;321-324,'X',329-333;334-339,
525;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <MCK>
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent comple
are linked by one or more disulfide bonds.
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a
C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciproca
inogen and may also play a role in the renin-angiotensin system by converting proreni

A:Gene: GDB:KLC3
A:Cross-references: GDB:127575; OMIM:229000
A:Map position: 4q35-4q35
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; in
F:1-10/Domain: signal sequence #status predicted <SIC>
F:20-638/Product: plasma kallikrein #status predicted <HCH>
F:20-638/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>

F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>
F:391-621/Domain: trypsin homology <TR>
F:21-104;47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,
F:127,308,396,453,494/Binding site: carboxylate (Asn) (covalent) #status experimenta
F:318-347,340-345/Disulfide bonds: #status predicted
F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 42.3%; Score 558.5; DB 1; Length 638;

[illegible]

RESULT 2

KORPUL

plasma kallikrein (EC 3.4.21.34) precursor - rat

N:Alternate names: Fletcher factor; kininogenin; serum kallikrein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Sep-1992 Sequence revision 30-Sep-1992 #text_change 18-Jun-1999

C:Accession: A39180, A33320, S06851, I53041, S06852

R:Beaubien, G.; Rosinaki-Chupin, I.; Mattei, M.G.; Mblakay, M.; Chretien, M.; Seldah, N.G.

R:Biochemistry 30, 1628-1635, 1991

A:Title: Gene structure and chromosomal localization of plasma kallikrein.

A:Reference number: A39180; MUID:91129236

A:Accession: A39180

A:Molecule type: DNA

A:Residues: 1-638 <BEA>

A:Cross-references: GB:J05315

A:Note: the authors translated the codon GAG for residue 81 as Gln

R:Seldah, N.G.; Lodenheim, R.; Mblakay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazure, N.G.

DNA 8, 563-574, 1989

A:Title: The cDNA structure of rat plasma kallikrein.

A:Reference number: A33320; MUID:90091743

A:Accession: A33320

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-638 <SEB>

A:Cross-references: GB:M30282; NID:g205010; PIDN:AAA4163.1; PID:g205011

A:Note: part of this sequence, including the amino ends of both the heavy and light chain

R:Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seldah, N.G.

Biochym. Biophys. Acta 999, 103-110, 1989

A:Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development

A:Reference number: S06851; MUID:90089457

A:Accession: S06851

A:Molecule type: protein

A:Residues: 20-45;391-413 <PAO>

R:Seldah, N.G.; Lodenheim, R.; Mblakay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazure, N.G.

DNA Cell Biol. 8, 563-574, 1989

A:Title: The cDNA structure of rat plasma kallikrein.

A:Reference number: I53041

A:Accession: I53041

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-638 <RES>

A:Cross-references: GB:M58590; NID:g206721; PIDN:AAA42069.1; PID:g206722

C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex with factor XIIa, which cleaves the molecule into a light and a heavy chain, which are linked by one or more disulfide bonds.

C:Genetics:

A:Gene: PK

A:Superfamily: coagulation factor XI; trypsin homology

C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla

F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
E:20-109/Domain: apple repeat <AP1>
E:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-264,227-256,231-237,292-375,318-347,
F:127,215,308,453,459,494/Binding site: carboxylate (Asn) (covalent) #status predicted
F:396/Binding site: carboxylate (Asp) (covalent) #status experimental
E:343,483,578/Active site: His, Asp, Ser #status predicted

	Query Match	Similarity	41.2%	Score	543:	DB	1:	Length	638;	
	Best Local	Similarity	42.9%	Pred.	No. 2e+41;					
	Matches	103:	Conservative	46:	Mismatches	85:	Indels	6:	Gaps	4:
QY	1	RIVGESALPGAMPQVSLHVNQV---	HVCGGSITPEPMIYAHCVCPLNNPWHMTAF	57		:		:		:
Db	390	PIRGVTNSSLGEMPRQVSLQVLYVSQNMMCGSSIIGRMILTAACFP-GIPDPVMRIY	448		:		:		:	
QY	58	AGILRFOSMFYTAGI-QVEKYISHPNYSKTKKNNDIALMKLQKPLETENDLYKPYCLPNPG	116		:		:		:	
Db	449	GGILMLSETTNKTSPSSIKELIIHQRYKMSSESYDIAIKOTPLANTYEFOPICLPESKA	508		:		:		:	
QY	117	NMLQPEOLCMIWISMGATEKEGTSEVLNAAKULILETORCSRYYVDLLIPAMICAFL	176		:		:		:	
Db	509	DNTNYLTNCWYTGMYTERGETONILOKATILPLYANECCAKTY-RDYVITRKMICACAYK	567		:		:		:	
QY	177	OGNVDSOQDSGGPLVYTSKNNIMWLIGDTSMWGSCAKARYPGVGYNVNPETMDIYROMRA	236	: :		: :		: :		
Db	568	EKGIDACKGSDSGPLVYCHSGSRMQLVGTISWGEGCARDEGCVYTKVAEYIDMWLEKIOS	627	: :		: :		: :		

```

RESULT      3
KOMSEPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: A36557
R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaublén, G.; Brachpapa, L.; Rochet
DNA CELL BIOL. 9, 737-748, 1990
A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compar
A:Reference number: A36557; MUID:91090844
A:Accession: A36557
A:Molecule type: mRNA
A:Residues: 1-638 <SE>
A:Cross-references: GB:M8588; NID:g200358; PIDN:AAA6393.1; PID:g200359
A:Note: part of this sequence, including the amino ends of both the heavy and light c
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent comple
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a
are linked by one or more disulfide bonds.
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; in
F:1-19/Domain: signal sequence #status predicted <IG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <LC>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,
F:127,215,308,396,494/Binding site: carboxylate (asn) (covalent) #status predicted
F:433,483,578/Active site: His, Asp, Ser #status predicted

```

Query Match 40.28; Score 530; DB 1; Length 638;
Best Local Similarity 42.58; Pred. No. 3e-40;
Matches 102; Conservative 43; Mismatches 89; Indels 6; Gaps 4;

THIS PAGE BLANK (USPTO)


```

FT DOMAIN 151 243 SRCR.
FT 256 492 CATALYTIC.
FT ACT_SITE 296 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 244 365 BY SIMILARITY.
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 465 BY SIMILARITY.
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 492 AA; 53847 MW; 7EAFEDF18609DDA CRC64;

Query Match 98.8%; Score 1303; DB 1; Length 492;
Best Local Similarity 98.3%; Pred. No. 3.6e-110;
Matches 234; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVNQVCGSITTEPMITVAHCVEKPLNPMHTAFAGI 60
DB 255 RIVGESALPGAMPQVSLHVNQVCGSITTEPMITVAHCVEKPLNPMHTAFAGI 314
QY 61 LROSFMEYAGAYQVQKVIKSHPNYDSKTKNDIALMKLQPLTFNDLVKPVCLPFGMMLQ 120
DB 315 LROSFMEYAGAYQVQKVIKSHPNYDSKTKNDIALMKLQPLTFNDLVKPVCLPFGMMLQ 374
QY 121 PEOLCMTISGKATEEKGKTSVLAQAKVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
DB 375 PEOLCMTISGKATEEKGKTSVLAQAKVLLIETORCNSRYVDNLITPAMICAGFLQGNV 434
QY 181 DSCGDSGGLPVTLSKNNIMWLIIGDTSMGSCAKAVRGVYGVNVVFTDWTYRQMRADG 238
DB 435 DSCGDSGGLPVTLSKNNIMWLIIGDTSMGSCAKAVRGVYGVNVVFTDWTYRQMRADG 492

RESULT 2
TMS2_MOUSE STANDARD; PRT; 490 AA.
ID TMS2_MOUSE
AC 09JIOB: 09JIC4: 09QY82;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-) (EPITHELIALIN) (PLASMIC
DE TRANSMEMBRANE PROTEIN X).
GN TMRPS2.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Vavaria M.H., Portvart K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT Expression of transmembrane serine protease TMRPS2 in mouse and
RT human tissues.*;
RL J. Pathol. 191:0-0(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT Putative transmembrane protease X.*;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT A novel mosaic serine protease, epithelisin.*;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.

```

```

CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
-----
DR EMBL: AF199362; AAF97867.1; -
DR EMBL: AF243500; AAF64186.1; -
DR EMBL: AF113596; AAF21308.1; -
DR HSSP: P00761; IAKS.
DR MCD: MGI:1354381; Tmrps2.
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -
DR InterPro: IPR002172; -
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS. 1.
DR PROSITE: PS00135; TRYPSIN_SER. 1.
DR PROSITE: PS01209; LDLRA_1. 1.
DR PROSITE: PS50068; LDLRA_2. 1.
KM Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
FT DOMAIN 1 83 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 84 104 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT 83 (POTENTIAL).
FT 104 EXTRACELLULAR (POTENTIAL).
FT 105 LDL-RECEPTOR CLASS A.
FT DOMAIN 111 149 SRCR.
FT DOMAIN 150 242 SRCR.
FT DOMAIN 254 490 CATALYTIC.
FT ACT_SITE 294 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 343 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 439 439 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 112 125 BY SIMILARITY.
FT DISULFID 119 138 BY SIMILARITY.
FT DISULFID 132 147 BY SIMILARITY.
FT DISULFID 243 363 BY SIMILARITY.
FT DISULFID 279 295 BY SIMILARITY.
FT DISULFID 408 424 BY SIMILARITY.
FT DISULFID 435 463 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 122 122 S->L (IN REF. 3).
FT CONFLICT 178 178 S->N (IN REF. 3).
FT CONFLICT 320 320 Y->H (IN REF. 1).
FT CONFLICT 474 474 N->D (IN REF. 1).
SQ SEQUENCE 490 AA; 53479 MW; 07DB03E4DBA1A9 CRC64;

Query Match 83.8%; Score 1105; DB 1; Length 490;
Best Local Similarity 84.4%; Pred. No. 2.6e-92;
Matches 200; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVNQVCGSITTEPMITVAHCVEKPLNPMHTAFAGI 60
DB 253 RIVGESALPGAMPQVSLHVNQVCGSITTEPMITVAHCVEKPLNPMHTAFAGI 312
QY 61 LROSFMEYAGAYQVQKVIKSHPNYDSKTKNDIALMKLQPLTFNDLVKPVCLPFGMMLQ 120
DB 313 LROSFMEYAGAYQVQKVIKSHPNYDSKTKNDIALMKLQPLTFNDLVKPVCLPFGMMLQ 372
QY 121 PEOLCMTISGKATEEKGKTSVLAQAKVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
DB 373 PEOLCMTISGKATEEKGKTSVLAQAKVLLIETORCNSRYVDNLITPAMICAGFLQGNV 432
QY 181 DSCGDSGGLPVTLSKNNIMWLIIGDTSMGSCAKAVRGVYGVNVVFTDWTYRQMRAD 237
DB 433 DSCGDSGGLPVTLSKNNIMWLIIGDTSMGSCAKAVRGVYGVNVVFTDWTYRQMRAD 489

```


RESULT	3
TMS3_HUMAN	STANDARD; PRT; 454 AA.
ID	TMS3_HUMAN
AC	P57727;
DT	01-OCT-2000 (Rel. 40, Created)
DT	01-OCT-2000 (Rel. 40, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DE	TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.-) (SERINE PROTEASE
DE	TADG-12) (TUMOR ASSOCIATED DIFFERENTIALLY-EXPRESSED GENE-12 PROTEIN).
GN	TPRPS3 OR TADG12 OR ECHOS1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCHI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED).
RC	TISSUE-Ovarian carcinoma;
RX	MEDLINE=20521358; PubMed=11068177;
RA	Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
RA	Wang Y., Parmley T.H., O'Brien T.J.?
RT	"Ovarian tumor cells express a novel multi-domain cell surface serine
RT	protease."
RL	Biochim. Biophys. Acta 1502:337-350(2000).
RN	[2]
RP	SEQUENCE FROM N.A. (ISOFORMS A; B/C AND D).
RX	PubMed=11137999;
RA	Scott H.S., Kudch J., Wattenhofer M., Shibuya K., Berry A., Chrast R.,
RA	Gyounis F., Mehdi S.O., Radhakrishna U., Papasavvas M.P., Gehrig C.,
RA	Rossler C., Korostitshevsky M., Gal A., Shimizu N., Bonne-Tamir B.,
RA	Anthonakis S.E.;
RT	"Insertion of beta-satellite repeats identifies a transmembrane
RT	protease causing both congenital and childhood onset autosomal
RT	recessive deafness.";
RL	Nat. Genet. 27:59-63(2001).
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC	-1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND
CC	TRUNCATED/TADG-1V; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED
CC	ISOFORM IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS.
CC	-1- DISEASE: DEFECTS IN TPRPS3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL
CC	NEUROSENSOR CHILDHOOD-ONSET FORMS OF DEAFNESS, DFNB8 AND DFNB10.
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC	TRYPSIN FAMILY.
CC	-1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC	-1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
DR	EMBL; AF201380; AAC37012.1; -
DR	EMBL; AB038157; BAB20077.1; -
DR	EMBL; AB038158; BAB20078.1; -
DR	EMBL; AB038159; BAB20079.1; -
DR	EMBL; AB038160; BAB20080.1; -
DR	MTM; 605511; -
DR	MTM; 601072; -
DR	MTM; 605316; -
DR	PROSITE; PS01209; LDLRA_1; 1.
DR	PROSITE; PS01209; LDLRA_2; 1.
DR	PROSITE; PS02068; LDLRA_2; 1.
DR	PROSITE; PSS0287; SRCR_2; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
KW	Hydrolase; Serine protease; Transmembrane; Signal-anchor; Deafness;
KW	Alternative splicing.
FT	DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 29 49 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT	(POTENTIAL).

[illegible]

RT contains four tandem repeats.";
 RL Biochemistry 25:2410-2417(1986).
 RN [2]
 RP PARTIAL SEQUENCE. AND DISULFIDE BONDS.
 RX MEDLINE-91152016; PubMed-1998666;
 RA McMullen B.A., Fujikawa K., Davie E.W.;
 RT "Location of the disulfide bonds in human plasma prekallikrein: the
 presence of four novel apple domains in the amino-terminal portion of
 the molecule."
 RL Biochemistry 30:2050-2056(1991).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
 HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
 THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- DISEASE: DEFECTS IN KLK3 ARE THE CAUSE OF FLETCHER FACTOR
 DEFICIENCY, A BLOOD COAGULATION DEFECT.
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPsin FAMILY. PLASMA KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-slb.ch/announce/>
 or send an email to license@isb-slb.ch).
 CC -----
 DR EMBL: M13143; AAA60153.1; -;
 DR PIR: A00921; KOHUP.
 DR PIR: A37939; A37939.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.212; -;
 DR MIM: 229000; -;
 DR InterPro: IPR000177; -;
 DR InterPro: IPR001254; -;
 DR InterPro: IPR001314; -;
 DR InterPro: IPR003014; -;
 DR Pfam: PF00024; PAN; 4.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PRO0005; APPLDOMAIN.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PROSITE: PS00495; APPL; 4.
 KW Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 KW Repeat.
 FT SIGNAL. 1 19
 FT CHAIN. 20 390
 FT CHAIN. 391 638 PLASMA KALLIKREIN HEAVY CHAIN.
 FT DOMAIN. 20 105 PLASMA KALLIKREIN LIGHT CHAIN.
 FT DOMAIN. 110 195 APPLE 1.
 FT DOMAIN. 200 285 APPLE 2.
 FT DOMAIN. 291 376 APPLE 3.
 FT DOMAIN. 389 621 APPLE 4.
 FT CARBOHYD. 127 127 CATALYTIC.
 FT CARBOHYD. 308 308 N-LINKED (GLCNAC. . .).
 FT CARBOHYD. 396 396 N-LINKED (GLCNAC. . .).
 FT CARBOHYD. 453 453 N-LINKED (GLCNAC. . .).
 FT CARBOHYD. 494 494 N-LINKED (GLCNAC. . .).
 FT ACCT_SITE. 483 483 CHARGE RELAY SYSTEM.
 FT ACCT_SITE. 483 483 CHARGE RELAY SYSTEM.
 FT ACCT_SITE. 578 578 CHARGE RELAY SYSTEM.
 FT DISULFID. 21 104
 FT DISULFID. 47 77
 FT DISULFID. 51 57
 FT DISULFID. 111 194

FT DISULFID 137 166
 FT DISULFID 141 147
 FT DISULFID 201 284
 FT DISULFID 227 256
 FT DISULFID 231 237
 FT DISULFID 292 375
 FT DISULFID 318 347
 FT DISULFID 322 328
 FT DISULFID 340 345
 FT DISULFID 383 503
 FT DISULFID 419 435
 FT DISULFID 517 584
 FT DISULFID 548 563
 FT DISULFID 574 602
 SQ SEQUENCE 638 AA; 71369 MW; E62F9C1053838F84 CRC64;
 Query Match 42.3%; Score 558.5; DB 1; Length 638;
 Best Local Similarity 43.4%; Pred. No. 6.4e-43;
 Matches 106; Conservative 46; Mismatches 83; Indels 9; Gaps 6;
 QY 1 RIVGESALPGAMPQVSLHQ--NVHYCGSITTEPMIVTAHCV- KPLNPFMHTA 56
 DB 390 RIVGNTSSWGEMPMQVSLQVAKTRQHRILCGSLIGHQWVLAHCFDGLPQDWR--I 447
 QY 57 FAGILROSEMFYAGY-QVEKVISHPNYSKTRNDIALMKLQKLPENDLVKPYCLPNP 115
 DB 448 YSGIILNSDITDFPFSQKEIITHQNVKVSCHNDIALIKLQAPLVNTEFGKPICLPSK 507
 QY 116 GMLPDLQWISGKATBEKKTSEVLNAKVLILEIQRCSKRVYDNLTPAMICAGF 175
 DB 508 GTSTIYNQWVTGWGFSKEKEIGNILQKVPILVTEECOKR-QDYKITQVWVACGY 566
 QY 176 LOGNVDSOCGSGPLPYTSKNINIMWLIGDTSWSSGCAKAYRPGVGNVFTWYRQMR 235
 DB 567 KRGKADACKGDSGRLVCKHNGMRLVGTISWGBECARREGQVTKYAEIWDWILEKTQ 626
 QY 236 -ADG 238
 DB 627 SSDG 630
 RESULT 5
 KAL_RAT
 ID KAL_RAT STANDARD; PRT; 638 AA.
 AC P14272;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN PK.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91129236; PubMed-1993180;
 RA Beaubien G., Rosinski-Chupin I., Mattei M.-G., Mbikay M., Chretien M.,
 Seidah N.G.;
 RT "Gene structure and chromosomal localization of plasma kallikrein.";
 RL Biochemistry 30:1628-1635(1991).
 [2]
 CC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE-90091743; PubMed-2598771;
 RA Seidah N.G., Ladenheim R., Mbikay M., Hamelin J., Lutfalla G.,
 Rougeon F., Lazure C., Chretien M.;
 RT "The cDNA structure of rat plasma kallikrein.";
 RL DNA 8:563-574(1989).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM

CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: M62357; AAA74563.1; -
 CC DR EMBL: M62358; AAA74563.1; JOINED.
 CC DR EMBL: M62346; AAA74563.1; JOINED.
 CC DR EMBL: M62347; AAA74563.1; JOINED.
 CC DR EMBL: M62349; AAA74563.1; JOINED.
 CC DR EMBL: M62350; AAA74563.1; JOINED.
 CC DR EMBL: M62351; AAA74563.1; JOINED.
 CC DR EMBL: M62352; AAA74563.1; JOINED.
 CC DR EMBL: M62353; AAA74563.1; JOINED.
 CC DR EMBL: M62354; AAA74563.1; JOINED.
 CC DR EMBL: M62355; AAA74563.1; JOINED.
 CC DR EMBL: M62356; AAA74563.1; JOINED.
 CC DR EMBL: M30282; AAA74563.1; -
 CC DR EMBL: M58590; AAA42069.1; -
 CC DR PIR: A39180; KORTPL.
 CC HSSP: P00750; IRTF.
 CC MEROPS: S01.212; -
 CC DR InterPro: IPR000177; -
 CC DR InterPro: IPR001254; -
 CC DR InterPro: IPR001314; -
 CC DR InterPro: IPR003014; -
 CC DR Pfam: PF00024; PAN; 4.
 CC Pfam: PF00089; Trypsin; 1.
 CC PRINTS: PR00005; APPLIEDOMAIN.
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 CC DR PROSITE: PS00135; TRYPSIN_SER; 1.
 CC DR PROSITE: PS00495; Apple; 4.
 CC Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 CC Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 CC Repeat.
 CC FT SIGNL. 1 19
 CC FT CHAIN 20 390
 CC FT CHAIN 391 638
 CC FT DOMAIN 20 105
 CC FT DOMAIN 110 195
 CC FT DOMAIN 200 285
 CC FT DOMAIN 291 376
 CC FT DOMAIN 389 621
 CC FT CARBOHYD 127 127
 CC FT CARBOHYD 308 308
 CC FT CARBOHYD 396 396
 CC FT CARBOHYD 453 453
 CC FT CARBOHYD 494 494
 CC FT ACT_SITE 434 434
 CC FT ACT_SITE 483 483
 CC FT ACT_SITE 483 483
 CC FT ACT_SITE 578 578
 CC FT DISULFID 21 104
 CC FT DISULFID 47 77
 CC FT DISULFID 51 57
 CC FT DISULFID 111 194
 CC FT DISULFID 137 166
 CC FT DISULFID 141 147
 CC FT DISULFID 201 284
 CC FT DISULFID 227 256

FT DISULFID 231 237 BY SIMILARITY.
 FT DISULFID 292 375 BY SIMILARITY.
 FT DISULFID 316 347 BY SIMILARITY.
 FT DISULFID 322 347 BY SIMILARITY.
 FT DISULFID 340 345 BY SIMILARITY.
 FT DISULFID 383 503 BY SIMILARITY.
 FT DISULFID 419 435 BY SIMILARITY.
 FT DISULFID 517 584 BY SIMILARITY.
 FT DISULFID 548 563 BY SIMILARITY.
 FT DISULFID 574 602 BY SIMILARITY.
 SQ SEQUENCE 638 AA; 71273 MW; 454DBE27E8CA8F88 CRC64;

 Query Match 41.2%; Score 543; DB 1; Length 638;
 Best Local Similarity 42.9%; Pred. No. 1,6e-41;
 Matches 103; Conservative 46; Mismatches 85; Indels 6; Gaps 4;

 QY 1 RIVGESALPGAMPQVSLHVN---HVGGSILTPPEWITVAHCEKPLNNPMTAF 57
 DB 390 RIVGNTSSIGEMPQVSLQVLSNHHGSGSILGRWILTAHCFD-GIPPDVWRIT 448
 QY 58 AGILROSEMFYAGY-OVERKIVSHPNYDSKTKNNDAIMKLRPLFNDLVKPCLPNPG 116
 DB 449 GGIILNSELTKPTFSSIKELIHOKYKMSGSDYDALIKLQPLNVTREFOKPCLPSRA 508
 QY 117 MMLQPELQWISGGAETEEKTSEVLNAAKVLLEFQRNSRVYVDNLITPAMICAGFL 176
 DB 509 DNTVITTCWVTGWTGKTEGTONILQKATPLVPEECOKRY-RDYVTTKOMICAGTK 567
 QY 177 QGNVSCQDSCGGLVTSKNNIMLIGDTSMGSCAKAPRGVYGNWVFTDVIYROMRA 236
 DB 568 EGGIDACKGDSGGPIVCKHSGRMQVIGITSMGSCARKRQPGYTVVAETIDILKIKS 627

 RESULT 6
 KAL_MOUSE STANDARD; PRT; 638 AA.
 AC P26262;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN KLK3 OR PK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN-BALB/C; TISSUE-Liver.
 RX MEDLINE=91090844; PubMed=2264928;
 RA Seidah N.G., Sawyer N., Hamelin J., Mion P., Beaubien G.,
 RA Brachpapa L., Rochemont J., Mblay M., Chretien M.,
 RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
 RT and comparison of protein and mRNA levels among species".
 RL DNA Cell Biol. 9:737-748(1990).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its


```

FT      DISULFID      416      432      CHAINS).
FT      DISULFID      514      581
FT      DISULFID      545      560
FT      DISULFID      571      599
FT      VARIANT      301      301      F -> L (IN FLI DEFICIENCY).
FT                                          /FtId-VAR.006622.
FT                                          C -> S (IN REF. 2).
FT CONFLICT      226      226
SQ SEQUENCE      625 AA; 70109 MW; 147AFA9AB7709EBF CRC64;

Query Match      40.1%; Score 529; DB 1; Length 625;
Best Local Similarity 42.2%; Pred. No. 2,9e-40;
Matches 103; Conservative 43; Mismatches 84; Indels 14; Gaps 5;

QY      1 RIVGESALPCAMFWQVSLHVN---HVGGSITTPMITYAANCVEKPLNPNHHTAF 57
      ||||| ::: : |||||::| : : ||||| : ||||| : :
Db      387 RIVGTSVREGEWPMQVLTHTTSPTRHLGCGSITIGNMWITLAHCF-YGVESPILRY 445

QY      58 AGILRSPM-----FYGAGYQVEKVISHPANDSKTKNNIDILMKLOKPLTENDLYKPYCL 112
      :||| :|| : :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      446 SGLINQSIRKEDTSFPG---VQELTIHDQYKMAESGIDILALKLETIVYNTDSORPCL 501

QY      113 PNPQMLLOPEOLCWSGNGATEEKGKTSEVLNAKVLLEIQRCSRYVDNLITPAMIC 172
      : : : : :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      502 PSKGDRLNYITDCWLTGNGYRKLRDKIQNTLOKAKIPLVTNECCQKR-RGHKITHKMK 560

QY      173 AGFLQGNVDSQGSNGSGPLVTSKNNIMWLIQDTISGSCACARVGVGNVAVFTDWIR 232
      || : : || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      561 AGYEGEGDACKGDSGGLSCKHNEVWHLVGLTSGEGCAQRERGVYTNVEYVDWILE 620

QY      233 QMRA 236
Db      621 KTAQ 624

RESULT      8
HEPS_HUMAN
ID      HEPS_HUMAN      STANDARD;      PRT;      417 AA.
AC      P05981;
DT      01-NOV-1988 (Rel. 09, Created)
DT      01-NOV-1988 (Rel. 09, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      SERINE PROTEASE HEPSIN (EC 3.4.21.-) (TRANSMEMBRANE PROTEASE, SERINE
      1).
OS      HPN OR TMPRSS1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Liver.
RX      MEDLINE=88209431; PubMed=2835076;
RA      Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davie E.W.;
RT      "A novel trypsin-like serine protease (hepsin) with a putative
RL      transmembrane domain expressed by human liver and hepatoma cells.";
RN      Biochemistry 27:1067-1074(1988).
RN      [2]
RP      CHARACTERIZATION.
RX      MEDLINE=91358502; PubMed=1885621;
RA      Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
RA      Chou S.H., Kurachi K.;
RT      "hepsin, a cell membrane-associated protease. Characterization,
RT      tissue distribution, and gene localization.";
RL      J. Biol. Chem. 266:16948-16953(1991).
RN      [3]
RP      CHARACTERIZATION.
RX      MEDLINE=93348237; PubMed=8346233;
RA      Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;
RT      "hepsin, a putative cell-surface serine protease, is a
RL      mammalian cell growth.";
RN      Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).

```

```

CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
CC OF CELL MORPHOLOGY.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES, WITH THE HIGHEST
CC LEVEL IN LIVER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M18930; AAA6013.1; -.
DR EMBL: X07732; CAA30558.1; -.
DR EMBL: X07002; CAA30058.1; -.
DR PIR: S00845; S00845.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.224; -.
DR MIM: 142440; -.
DR InterPro: IPR001354; -.
DR InterPro: IPR001314; -.
DR Pfam: PF00089; Trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 162 NON-CATALYTIC CHAIN (POTENTIAL).
FT CHAIN 163 417 CATALYTIC CHAIN (POTENTIAL).
FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 18 44 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 45 417 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 203 203 CATALYTIC.
FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 353 353 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 153 277 INTERCHAIN (BY SIMILARITY).
FT DISULFID 188 204 BY SIMILARITY.
FT DISULFID 322 338 BY SIMILARITY.
FT DISULFID 349 381 BY SIMILARITY.
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 417 AA; 45011 MW; B2086FF61E51D7 CRC64;

```

```

Query Match 40.1%; Score 528.5; DB 1; Length 417;
Best Local Similarity 41.2%; Pred. No. 2e-40; Indels 11; Gaps 3;
Matches 101; Conservative 41; Mismatches 92;

```

```

OY 1 RIVGESALPGAMPQVSLAHVONVHVGSGSITPEWITVAACVKEPLNNPMHTAFAGI 60
DB 169 RIVGSRDRLSLGMPQVSLAHVONVHVGSGSITPEWITVAACVKEPLNNPMHTAFAGI 221
OY 61 LROSTMEFGAGYQVEKVSHPNY-----DSKTKNDALMLKQPLTFENDLVKPCVLEP 114
DB 222 VAOASR-HGLQGVAVVYHGGYLPFRDPNSSENSNDIALVHLSPLPLEYIQPCVLA 280
OY 115 PCMMIQLPQLCISGMGATTEKGTSEVLANAKVLLIEFRQNSRKYVONLTPMATICG 174
DB 281 AGQALVLDGICITVGTWGNTOYVGOAGVQLQEARVPLISDVNGADFGNOIKPKMFCAG 340
OY 175 FLQGVNDSQCGSGGSPV---TSKNINIMLIGDTSWGSAGCAKAYRPGYGVNMYFTDWI 230
DB 341 YREGIDACQSGSGGSPFCEDSISRTPRRLCGIYSGWIGCALQKPGVYTVSPDFEWI 400
OY 231 YRQMR 235
DB 401 FOAIK 405

```

```

RESULT 9
ENTK_BOVIN
AC P98072;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRS7 OR ENTK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Duodenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
[2]
RN SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043122; PubMed=8226855;
RA Lavallie E.R., Rehmentulla A., Racine L.A., Dblasio E.A.,
RA Fernandez C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase.";
RT J. Biol. Chem. 268:23311-23317(1993).
[3]
RN SEQUENCE OF 801-827.
RC TISSUE=Intestine;
RX MEDLINE=92189715; PubMed=1799406;
RA Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinase.";
RT J. Protein Chem. 10:475-480(1991).
[4]
RN -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
RN PROTEOLYTIC PROENZYMES (TRYPsin, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
RN A). IT CATALYZES THE CONVERSION OF TRYPsin TO TRYPsin WHICH IN
RN TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
RN PROCARBOXYPEPTIDASES, AND PROELASTASES.
RN -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
RN TRYPsinOGEN.
RN -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
RN MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
RN -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
RN -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
RN -1- CLEAVED BY A TRYPsin-LIKE PROTEASE.
RN -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
RN -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
RN -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
RN -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
RN -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
RN TRYPsin FAMILY.
RN -----
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN -----
DR EMBL: U09859; AAA40026.1; -.
DR EMBL: L19663; AAA16035.1; -.
DR PIR: A61436; A61436.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.156; -.
DR InterPro: IPR000082; -.
DR InterPro: IPR000859; -.

```



```

RESULT 11
HEPS_MOUSE
ID HEP5_MOUSE STANDARD: PRT: 416 AA.
AC 035453;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SERINE PROTEASE HEP5IN (EC 3.4.21.-).
GN HPN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver; PubMed=9395459;
RA MEDLINE=98058912; Haakma C., Tomasek J.J., Howard E.W.;
RA Vu T.-K.H., Liu R.W.,
RT "Identification and cloning of the membrane-associated serine
RT protease, hepsin, from mouse preimplantation embryos."
RT J. Biol. Chem. 272:31315-31320(1997).
RL J. Biol. Chem. 272:31315-31320(1997).
CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
CC OF CELL MORPHOLOGY.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF030065; AAB84221.1; -.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.224; -.
DR MGD: MGI:1196620; Hpn.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR Pfam: PF00089; Trypsin; 1.
DR PRINTS: PRO0722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR KEGG: Hydroxylase; Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 161 NON-CATALYTIC CHAIN (POTENTIAL).
FT CHAIN 162 416 CATALYTIC CHAIN (POTENTIAL).
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 44 416 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 162 416 CATALYTIC.
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 256 256 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 352 352 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 152 276 INTERCHAIN (BY SIMILARITY).
FT DISULFID 187 203 BY SIMILARITY.
FT DISULFID 321 337 BY SIMILARITY.
FT DISULFID 348 380 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 416 AA; 44739 MW; 432194FF4004F848 CRC64;

```

Query Match 38.5%; Score 507.5; DB 1; Length 416;
 Best Local Similarity 40.4%; Pred. No. 1.6e-38;
 Matches 99; Conservative 42; Mismatches 93; Indels 11; Gaps 3;

```

QY 1 RIVGSALEPGAPQVSLVNVAVHVGSGSIITPEIVTAHACVEKPLNPMHTAFAGI 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 161 RIVGQDSSIGRPMQVSLRYDGTSLHCGSLSGDWVLTAAHCFPEKNRRLSKRRVAGA 220
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 61 LRQSFMEYGAGYOVERKVISHPN-----DSKTRKNDIALMKLQKPLTFNDLVKPYCLPN 114

```

```

RESULT 12
ENTR_MOUSE
ID ENTR_MOUSE STANDARD: PRT: 1069 AA.
AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
GN PRS7 OR ENTR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Duodenum;
RA MEDLINE=98147142; PubMed=9486188;
RA Yuan X., Zheng X., Lu D., Rudin D.C., Pung C.Y.M., Sadler J.E.;
RT Structure of murine enterokinase (enteropeptidase) and expression in
RT small intestine during development."
RT Am. J. Physiol. 274:G342-G349(1998).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U73378; AAB3717.1; -.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.156; -.
DR MGD: MGI:1197523; Prs7.
DR InterPro: IPR000082; -.
DR InterPro: IPR000859; -.
DR InterPro: IPR000988; -.
DR InterPro: IPR001190; -.

```

```

DR InterPro; IPR001254; -
DR InterPro; IPR001314; -
DR InterPro; IPR002172; -
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF00390; SEA; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00057; IGL_recept.a; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS00740; MAM; 1.
DR PROSITE; PS00060; MAM; 2.
DR PROSITE; PS01209; LDLRA; 1.
DR PROSITE; PS00068; LDLRA; 2.
KW Signal-anchor: Glycoprotein; Myristate; Hydrolase;
KM Serine protease; zymogen; Transmembrane; Repeat.
FT CHAIN 1 829 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT TRANSMEM 19 47 CATALYTIC CHAIN (LIGHT CHAIN).
FT DOMAIN 227 268 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 270 379 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 387 549 CUB.
FT DOMAIN 569 679 MAM.
FT DOMAIN 686 724 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 723 816 SRCR.
FT ACT_SITE 874 874 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 925 925 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1021 1021 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LIPID 2 2 MYRISTATE (POTENTIAL).
FT DISULFID 229 242 BY SIMILARITY.
FT DISULFID 236 255 BY SIMILARITY.
FT DISULFID 249 266 BY SIMILARITY.
FT DISULFID 688 700 BY SIMILARITY.
FT DISULFID 695 713 BY SIMILARITY.
FT DISULFID 707 722 BY SIMILARITY.
FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
FT DISULFID 859 1027 BY SIMILARITY.
FT DISULFID 959 1027 BY SIMILARITY.
FT DISULFID 991 1006 BY SIMILARITY.
FT DISULFID 1017 1045 BY SIMILARITY.
FT CARBOHYD 147 147 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 579 579 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 675 675 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 727 727 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 770 770 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 791 791 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAG. . .) (POTENTIAL).
FT CARBOHYD 999 999 N-LINKED (GLCNAG. . .) (POTENTIAL).
SO SEQUENCE 1069 AA; 118735 MW; E62549E463743C3D CRC64;

Query Match 38.2%; Score 504; DB 1; Length 1069;
Best Local Similarity 41.2%; Pred. No. 9.3e-38;
Matches 98; Conservative 50; Mismatches 82; Indels 8; Gaps 4;

```

```

OY 1 RYGGESALPGAMWQVSLHGVNH-----VCGSGITTPREMITAACHVEKPLANNPMHTA 56
DB 829 RYGGESDQAGAMPVVALYHHRDSTDRLLCGASLVSSDWLVAACHYRRNLDPTRMTA 888
OY 57 FAGLROSEFMFYAGYQ--VEKVISHPNVDSKTKNNNDLMLKQPLTFNDVKPVLCPN 114
DB 889 VGLGLMQSLNLSPOVYVRVVDQIVINPHYDRRKRYNDIAMMLERKVVNTDYIQICLPE 948

OY 115 PGMMLPEQLCISGWGATE-EKGTSEVLNNAKXLLLETPCRNSRYVNDLITPAMICA 173
DB 949 ENQIFIPGRGCSLACGKXINAGSVVDLKEADYPLISNEKCCQQLPEYN-ITSMICA 1007
OY 174 GFLQGVNDSGCGDSGPLYTSKNNIMWLIGDTSGSGCAKATRPQYGVNMYFTWIT 231
DB 1008 GYEEGIDSCQDSGSGPLMCCENNRWFLGVTSFGVOCALPNHPGVYRVVSQFIEMH 1065

RESULT 13
PLMN_MOUSE STANDARD; PRT; 812 AA.
ID PLMN_MOUSE
AC P20918;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLASMINOGEN PRECURSOR (EC 3.4.21.7) [CONTAINS: ANGIOSTATIN].
GN PLG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91184812; PubMed=2081600;
RA Degen S.J., Bell S.M., Schaefer L.A., Elliott R.W.;
RT "Characterization of the cDNA coding for mouse plasminogen and
RL localization of the gene to mouse chromosome 17.";
RN Genomics 8:49-61(1990).
RP CHARACTERIZATION OF ANGIOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE-95042728; PubMed=7525077;
RA O'Reilly M.S., Holmgren L., Shing Y., Chen C., Rosenthal R.A.,
RA Moses M., Lane W.S., Cao Y., Sage E.H., Folkman J.;
RT "Angiostatin: a novel angiogenesis inhibitor that mediates the
RL suppression of metastases by a Lewis lung carcinoma.";
RL Cell 79:315-328(1994).
CC -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC AND INFLAMMATION. IN OVULATION IT WEAKENS THE WALLS OF THE
CC GRAFAN FOLLICLE. IT ACTIVATES THE UKONINASE-TYPE PLASMINOGEN
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC LAMININ AND VON WILLEBRAND FACTOR.
CC -1- FUNCTION: ANGIOSTATIN IS AN ANGIOGENESIS INHIBITOR THAT BLOCKS
CC NEOVASCULARIZATION AND GROWTH OF EXPERIMENTAL PRIMARY AND
CC METASTATIC TUMORS IN VIVO.
CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC FIBRIN. CANNOT BE ACTIVATED WITH STEPTOKINASE.
CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
CC -1- MISCELLANEOUS: IN THE PRESENCE OF THE INHIBITOR, THE ACTIVATION
CC INVOLVES ONLY CLEAVAGE AFTER ARG-581, RESULTING IN 2 CHAINS HELD
CC TOGETHER BY 2 DISULFIDE BONDS. WITHOUT THE INHIBITOR, THE
CC ACTIVATION INVOLVES ALSO REMOVAL OF THE ACTIVATION PEPTIDE.
CC -1- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMINOGEN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: J04766; AAA50168.1; -
CC PIR: A38514; A38514.
CC HSPB; P00747; 1PKK.

```

[illegible]

```
Query Match Score 38.1%; Seq ID 50.2; DB 1; Length 812;  
Best Local Similarity 43.3%; Pred. No. 9,4e-38;  
Matches 106; Conservative 36; Mismatches 78; Indels 25; Gaps 7;  
  
OY      1 RIVGESALPGAMPQVSLHVO--NVHVCGSGIITPEPVIATAHCEKPLNNPMHTAF 58  
       |::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db      561 RVVGCCVANPHSMWPIQISLRTFTGQHCGGLIPAEWVLTAAHCLERS-SRPEFYKVL 639  
  
OY      59 GILROSFMEYGAGYOVE-----KVISHPNYSKTNNDIALMLCKPLFENDLVKPVCLP 113  
       ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db      640 GAHEE-----YINGLDVQELSVAKLLPER-----NRDIAALLKLSPNPTTKVIPACLP 689  
  
OY     114 NGMMLOPQLCWMISGWATE--EKGTSEVLANAKVLLLETFQRNSRYVDNLITPM 170
```

```

Db      690 SPNVYVADRTICITCYTGMKETOCTGFACGLAKE-----AQLPVLENKACNRYELINRVKSTTE 745
QY      171 ICAGFLAGNVDSOCQSDSGGPIVLTSSKNNTWMLIGDTSSGSCAKAYRPGVYGVVFTDMI 230
Db      746 LCAGGLAGGVDSQCQSDSGPLVCFEKDKYILIGVYTWGLGCGARPRKPGVYRVSRFVDMI 805
QY      231 YRQMR 235
Db      806 EREMR 810

RESULT 14
ST14_HUMAN STANDARD: PRT: 855 AA.
ID ST14_HUMAN
AC Q9Y5Y6;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SDRPRESSOR OF TUMORICENTICITY 14 (EC 3.4.21.-) (MATRIPTASE) (MEMBRANE-
DE TYPE SERINE PROTEASE 1) (MT-SP1).
GN ST14 OR PRSS14 OR SNCL9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99303581; PubMed=10373424;
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matriptase, a matrix-degrading serine
RT protease with trypsin-like activity."
RN J. Biol. Chem. 274:18231-18236(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Takeuchi T., Shuman M.A., Craik C.S.;
RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
RT dissect complex biological processes and identify a membrane-type
RT serine protease in epithelial cancer and normal tissue."
RN Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RN [3]
RP CHARACTERIZATION.
RA TISSUE=Milk;
RX PubMed=10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matriptase
RT and a Kunitz-type serine protease inhibitor from human milk."
RN J. Biol. Chem. 274:18237-18242(1999).
CC -1- FUNCTION: DEREGATES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
CC OR LYS AS THE P1 SITE.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: CONTAINS 4 LDI-RECEPTOR CLASS 4 DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF118224; AAD42765.2; -
DR EMBL, AF133086; AAF00109.1; -
DR HSSP: P00763; IDPO.
DR MEROPS: S01.302; -
DR InterPro: IPR000859; -
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -

```


Wed Sep 26 13:57:02 2001

us-09-615-285-2_copy_255_492.rsp

Page 15

Search completed: September 26, 2001, 09:27:27
Job time: 317 sec

2001-09-26 13:57:02

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:27:03 : Search time 40.65 Seconds
(without alignments)
774.627 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492

Perfect score: 1319

Sequence: 1 RIVGESALPGAMPWQVSLH.....VYGNVWFTDWIRQMRADG 238

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP-REMBL_16: *
2: sp-archaea: *
3: sp-bacteria: *
4: sp-fungi: *
5: sp-human: *
6: sp-invertebrate: *
7: sp-mammal: *
8: sp-mhc: *
9: sp-organellar: *
10: sp-phage: *
11: sp-plant: *
12: sp-rodent: *
13: sp-unclassified: *
14: sp-vertebrate: *
15: sp-virus: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	705	53.4	767	13	Q9DGR2
2	558	42.3	457	4	Q9H3S3
3	554	42.0	643	6	O97506
4	549	41.6	273	11	O9ER02
5	549	41.6	311	11	O9ER04
6	548	41.5	445	11	O9ER03
7	529	40.1	571	4	O9Y495
8	524.5	39.8	310	11	O9OY29
9	521	39.5	423	4	O9N2A5
10	521	39.5	437	4	O9NRS4
11	518.5	39.3	418	4	O60235
12	504	38.2	855	11	O9J117
13	502	38.1	855	4	O9HCA3
14	499	37.8	1524	13	O91674
15	498	37.8	855	4	O9H350
16	497.5	37.7	279	11	O9OZ74
17	497	37.7	389	13	O9PVX7
18	496	37.6	305	11	O9JHJ7
19	493.5	37.4	812	11	O9ROW3

20	492.5	37.3	422	4	O9UL52
21	490.5	37.2	329	6	O9GL10
22	481.5	36.5	415	6	O29015
23	479	36.3	845	13	O9DGR1
24	465.5	35.3	297	11	O88781
25	465.5	35.3	312	4	O9NS34
26	464.5	35.2	421	11	O60491
27	464	35.2	1042	4	O9Y505
28	459	34.8	273	6	O9XSM2
29	458.5	34.8	311	11	O9OUL7
30	458	34.7	307	4	O9OHY2
31	457.5	34.7	377	6	P79343
32	455.5	34.5	263	5	O62562
33	455.5	34.5	271	4	O43342
34	455.5	34.5	314	4	O9Y6M0
35	455.5	34.5	317	4	O9GZM4
36	455.5	34.5	321	4	O9NR08
37	455	34.5	306	11	O9ER10
38	454.5	34.5	321	4	O9UBB2
39	452	34.3	342	11	O9ES87
40	452	34.3	342	11	O9ER01
41	445.5	33.8	317	13	O9DGR3
42	445	33.7	733	5	O9VFX9
43	442.5	33.5	263	5	O9TY16
44	442.5	33.5	266	5	O27761
45	442	33.5	334	6	O46507

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	767 AA.
Q9DGR2				
AC	O9DGR2;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	EMBRYONIC SERINE PROTEASE-2.			
GN	XESP-2.			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;			
OC	Xenopodinae; Xenopus.			
OX	NCBI-TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20363741; PubMed=10903452;			
RA	Yamada K., Takabatake T., Takeshima K.;			
RT	"Isolation and characterization of three novel serine protease genes			
RT	from Xenopus laevis."			
RL	Gene 252:209-216(2000).			
DR	EMBL; AB038497; BAB08217.1; -.			
KW	Protease.			
SQ	SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;			

Query Match 53.4%; Score 705; DB 13; Length 767;
Best Local Similarity 54.9%; Pred. No. 5.3e-63;
Matches 129; Conservative 29; Mismatches 77; Indels 0; Gaps 0;

QY	1	RIVGESALPGAMPWQVSLHVNWHVCGSITPPEVTFATGAVKRPINPMHTAFAGI 60
DB	530	RIVGGTFANLGNMPWQVNLQITGVLCGGSITSPKWLVTATACVYSSASASMRVPAGT 589
QY	61	LROSFMEYAGYOVKEYISHPNYDSKTKNNDIALMKLQKLPFNLDLVKPYCLPDPGMMQ 130
DB	590	LTRPSYNNASAVYERLIVHPRKSYTYDDIALMKLRDEITGYTTPQPCPLNSGFMWE 649
QY	121	PEQLCITSGGATDEKGTSEVLNAKVLLETRCKSRIVYDNLITPATICGFLQGNV 180
DB	650	AGTTTWISGWSYEGSVSTYLYAFAIPLDSNVCQSVYNGOITSSMICAGYLSGV 709

Oy	181	DSCGGDGGGLVATSKNNIMWLIGDTSMSGCGAKAARPCPVGNWVFWDMYRQMR	235
	:	: : :: :	
Db	710	DTCCGDSGGFLVKNKGTVMVLVGDTSMGDGCARANKPGCVYGNVTTFLEWTYSQMR	764
RESULT	2		
09H3S3			
ID	09H3S3	PRELIMINARY;	PRT: 457 AA.
AC	09H3S3;		
DT	01-MAR-2001 (Tremblrel, 16,	Created)	
DT	01-MAR-2001 (Tremblrel, 16,	Last sequence update)	
DE	SPINESIN.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN PROSTATE;		
RA	Mitsui S., Yamaguchi N.;		
RT	"Molecular cloning of human spinesin.";		
RL	Submitted (MAY-1999) to the EMBL/Genbank/DDBJ databases.		
DR	EMBL: AB028140; BAB20375.1; -		
SO	SEQUENCE	457 AA; 49574 MW; 64406BAR98542651 CRC64;	

Query Match	42.3%	Score 558	DB 4	Length 457
Best Local Similarity	45.1%	Pred. No. 2.3e+48		
Matches 105	Conservative 31	Mismatches 95	Indels 2	Gaps 2

[illegible]

```

RESULT      3
097506
ID      097506      PRELIMINARY;      PRT;      643 AA.
AC      097506;
DT      01-MAY-1999 (TReMBLrel. 10, Created)
DT      01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT      01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE      KALLIKREIN.
OS      Sus scrofa (Pig) .
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RX      NCBI_TaxId=9823;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=PLASMA;
RA      Tekanashi T., Kimura A., Okimura H., Hamabata T.;
RT      "Porcine liver plasma kallikrein.";
RL      Submitted (JAN-1999) to the EMBL/GenBank/DBD databases.
CC      -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC      -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (SI) .
DR      EMBL; AB022425; BAA3147.1; -.
DR      HSSP; P00766; 1CHG.
DR      MEROPS; S01.212; -.
DR      InterPro; IPR000177; -.
DR      InterPro; IPR001254; -.

```

DR	InterPro: IPR001314; -	
DR	InterPro: IPR003014; -	
DR	Pfam: PF00024; PAN; 4.	
DR	Pfam: PF00089; trypsin; 1.	
DR	PRINTS: PR00005; APPLIEDMAIN.	
DR	PRINTS: PR00722; CHYMOTRYPSIN.	
DR	PROSITE: PS00495; APPLE; 4.	
DR	PROSITE: PS00134; TRYPsin_HIS; UNKNOWN_1.	
DR	PROSITE: PS00135; TRYPsin_SER; UNKNOWN_1.	
KM	SMART: SM00020; Tryp_Spc; 1.	
KM	Hydrolase: Serine protease.	
SQ	SEQUENCE: 643 AA; 72227 MW; AAF2933E3C3CB80A CRC64;	

Query Match	42.0%;	Score 554;	DB 6;	Length 643;
Best Local Similarity	44.2%;	Pred. No. 9e-48;		
Matches 107;	Conservative 40;	Mismatches 87;	Indels 8;	Gaps 5

```

0Y      1 RIVGE$ALPGAMPOVS-----LHVQVHVGGSITRPENVTAAHCEVPLNNPMTA 56
Db      398 RIVG$TDFLEGMPOVSLQAKLRQN-HLGG$SIIGHQVUTAAHCFD-G$ISLPQMRI 455$
0Y      57 FAGILRO$FMFGAGY-QEYK$ISHPNTD$TKNNIDALMLQ$KPLTFNDLVK$CLPNP 115$
Db      456 YGGLINISEITREPE$OVQK$EILIHQNKILIESCHDIALKL$PLPNTYTFQ$RIPC$R 515$

```

```

      11: 111 1111: 1 1: 1 1: 1111:
516 DDINVVYTCWATGWMGFTTEEGEIQNILQKVNIPVLSNECQKSY - RDHKISKOMICAGY 574
Db

```

```

Db      575 KEGGKDACGESGGPLVCKYNGIWMHVLVGTTSMWEGECARREQPGVYTKVLEYMDWILEKTQ 634
          :| |:|:|:||||| | || |:| ||| |:| |:| |:| |:| :|

```

Db 635 DP 636

RESULT	4
Q9ER02	
ID	Q9ER02
PRELIMINARY;	
PRT;	273 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE type 1 SPINESIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Mitsui S., Yamaguchi N.;
RT "cDNA cloning of mouse type 1 spinesin.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
EMBL AB016423; BAB20278.1, -
SQ SEQUENCE 273 AA; 29662 MW; 63E45FAD2BDF75D CRC64;

Query Match	41.68;	Score 549;	DB 11;	Length 273;
Best Local Similarity	42.98;	Pred. No. 9.9e-48;		
Matches 100;	Conservative 37;	Mismatches 94;	Indels 2;	Gaps 2;

0Y 1 RIVGESALPAMPOVSLHNOAHVGGGSIITPEWITVAHCYEK - PLNNPNHHNPAFG 59
| | | | | : | | | | : | | | | : | | | | : | |
Db 35 RIVGGAAASGRMWRQASVMLSGRHTGCASVLAPHNWVTVAHCMTSRRLSRLSSWKRYAG 94
| | | | | : | | | | : | | | | : | | | | : | |
0Y 60 ILROSEFMYGAGOVEKYISHPNTDSTKNNDIALMLQKPPLTFNDLVKRECVLPNGMIL 119
| | | | | : | | | | : | | | | : | | | | : | |
Db 95 LVSHCAVQHOGTWEKEIIPRLYSQNHBDVALDLQRPINSDVDVAACPRAEQYE 154
| | | | | : | | | | : | | | | : | | | | : | |

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:22:08 : Search time 20.35 Seconds
(without alignments)
240.811 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492
Perfect score: 1319
Sequence: 1 RIVGESALPGAMPWQVSLH.....VYGNVMVFTDWTYRQNRADG 238

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCCTS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1319	100.0	492	4	US-09-342-749-2 Sequence 2, Appl1
2	1310	99.3	283	3	US-08-807-151-1 Sequence 1, Appl1
3	553.5	42.0	248	4	US-08-944-483-63 Sequence 63, Appl1
4	543	41.2	638	2	US-08-681-151-3 Sequence 3, Appl1
5	528.5	40.1	256	2	US-09-027-337-3 Sequence 3, Appl1
6	524	39.7	238	4	US-08-944-483-64 Sequence 64, Appl1
7	523.5	39.7	255	4	US-08-944-483-67 Sequence 67, Appl1
8	523.5	39.7	798	5	US-08-200-900A-2 Sequence 2, Appl1
9	523.5	39.7	798	5	PCR-US94-00616-2 Sequence 2, Appl1
10	521.5	39.5	235	3	US-08-807-151-3 Sequence 3, Appl1
11	521	39.5	435	4	US-09-008-271A-6 Sequence 6, Appl1
12	518.5	39.3	418	1	US-08-508-448C-25 Sequence 25, Appl1
13	513.5	38.9	232	1	US-08-508-448C-19 Sequence 19, Appl1
14	511.5	38.8	235	4	US-08-944-483-65 Sequence 65, Appl1
15	507.5	38.5	416	2	US-09-000-846-2 Sequence 2, Appl1
16	502.5	38.1	812	1	US-08-248-629A-1 Sequence 1, Appl1
17	502.5	38.1	812	1	US-08-451-932-1 Sequence 1, Appl1
18	502.5	38.1	812	1	US-08-452-260-1 Sequence 1, Appl1
19	502.5	38.1	812	1	US-08-326-785-1 Sequence 1, Appl1
20	502.5	38.1	812	2	US-08-612-788-1 Sequence 1, Appl1
21	502.5	38.1	812	2	US-08-605-598B-1 Sequence 1, Appl1
22	502.5	38.1	812	2	US-08-429-743-1 Sequence 1, Appl1
23	502.5	38.1	812	2	US-08-866-735-1 Sequence 1, Appl1
24	502.5	38.1	812	3	US-09-066-028-1 Sequence 1, Appl1
25	502.5	38.1	812	5	PCR-US95-05107-1 Sequence 1, Appl1
26	502	38.1	855	2	US-09-027-337-2 Sequence 2, Appl1
27	495.5	37.6	250	4	US-08-944-483-68 Sequence 68, Appl1

28	459	34.8	356	2	US-08-681-151-1 Sequence 1, Appl1
29	455.5	34.5	314	4	US-09-008-271A-3 Sequence 3, Appl1
30	450	34.1	270	2	US-08-978-404B-8 Sequence 8, Appl1
31	447	33.9	276	2	US-09-016-366A-15 Sequence 15, Appl1
32	447	33.9	276	2	US-08-978-404B-21 Sequence 21, Appl1
33	446	33.8	230	1	US-08-379-621-2 Sequence 2, Appl1
34	446	33.8	230	1	US-08-147-000B-2 Sequence 2, Appl1
35	446	33.8	230	2	US-08-889-078-2 Sequence 2, Appl1
36	444	33.7	299	4	US-08-944-483-66 Sequence 66, Appl1
37	443	33.6	791	1	US-08-643-219-1 Sequence 1, Appl1
38	443	33.6	791	1	US-08-851-350-1 Sequence 1, Appl1
39	442.5	33.5	221	4	US-08-944-483-54 Sequence 54, Appl1
40	442.5	33.5	222	1	US-08-456-840-46 Sequence 46, Appl1
41	442.5	33.5	222	1	US-08-266-407A-46 Sequence 46, Appl1
42	442.5	33.5	222	2	US-08-892-544-46 Sequence 46, Appl1
43	442	33.5	546	6	5200340-6 Patent No. 5200340
44	442	33.5	790	1	US-08-469-486-54 Sequence 54, Appl1
45	442	33.5	790	2	US-08-469-658-54 Sequence 54, Appl1

ALIGNMENTS

RESULT 1
US-09-342-749-2
Sequence 2, Application US/09342749
Patent No. 6166194
GENERAL INFORMATION:
APPLICANT: Mong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TMPPRS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-342-749-2

Query Match 100.0%; Score 1319; DB 4; Length 492;
Best Local Similarity 100.0%; Pred. No. 8.7e-139;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPWQVSLHVNWHVCGSIITPEWIVTAHCEVKEPLNNPMTAFAGI 60
|||||
DB 255 RIVGESALPGAMPWQVSLHVNWHVCGSIITPEWIVTAHCEVKEPLNNPMTAFAGI 314
|||||
QY 61 LROSFMEYAGYQVEVYISHPNVDSKTKNDIALMLQKPLTFNDLVKPCLPNPMQLQ 120
|||||
DB 315 LROSFMEYAGYQVEVYISHPNVDSKTKNDIALMLQKPLTFNDLVKPCLPNPMQLQ 374
|||||
QY 121 PEOLCMTSGGATEEKGKTESEVLAIAKVLIIETORNSRYVDNLITPAMICAGFLQGNV 180
|||||
DB 375 PEOLCMTSGGATEEKGKTESEVLAIAKVLIIETORNSRYVDNLITPAMICAGFLQGNV 434
|||||
QY 181 DSCQGSGLPVTYSKNNIMLIDITSGSCCAKAYRPGVGNVFTDWTYRQNRADG 238
|||||
DB 435 DSCQGSGLPVTYSKNNIMLIDITSGSCCAKAYRPGVGNVFTDWTYRQNRADG 492
|||||

RESULT 2
US-08-807-151-1
Sequence 1, Application US/08807151
Patent No. 6043033
GENERAL INFORMATION:

```
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
TITLE OF INVENTION: PROTEASE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Inocyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807,151
FILING DATE: Filed Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0227 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCORNO701
CLONE: 556016
US-08-807-151-1

Query Match          99.3%; Score 1310; DB 3; Length 283;
Best Local Similarity 99.6%; Pred. No. 3,9e-138;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPQVSLHYQNVHVCSSITTPMIVTAHCVCKPLNNPWHMTAFAGI 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 46 RIVGESALPGAMPQVSLHYQNVHVCSSITTPMIVTAHCVCKPLNNPWHMTAFAGI 105
OY 61 LRQSFMEFYAGAYQVEKVISHPNYSKTKNNDIAMKLRKPLTFNDLVKPVCLPNFGMLQ 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 106 LRQSFMEFYAGAYQVEKVISHPNYSKTKNNDIAMKLRKPLTFNDLVKPVCLPNFGMLQ 165
OY 121 PEOLCWSISGMKATEKTKTSEVLNAKAVLLLETORCNSRYVYDNLITPAMICAGFLQGNV 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 166 PEOLCWSISGMKATEKTKTSEVLNAKAVLLLETORCNSRYVYDNLITPAMICAGFLQGNV 225
OY 181 DSCGDSGGLPVTSKNNIMWLLIGDTSWGSCKAKAVRPGVYGVNVVFTDWTYRQMRADG 238
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 226 DSCGDSGGLPVTSKNNIMWLLIGDTSWGSCKAKAVRPGVYGVNVVFTDWTYRQMRADG 283

RESULT 3
US-08-944-483-63
; Sequence 63; Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
```

```
APPLICANT: KLAAS, MICHAEL R.
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STEWART, KENT D.
APPLICANT: STROUPE, STEVEN D.
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESS: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,483
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6183-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6232456e
US-08-944-483-63

Query Match          42.0%; Score 553.5; DB 4; Length 248;
Best Local Similarity 43.2%; Pred. No. 8.4e-54;
Matches 105; Conservative 46; Mismatches 83; Indels 9; Gaps 6;

OY 2 IYGGESALPGAMPQVSLHYQ---NVHVCSSITTPMIVTAHCVK-PLNNPWHMTAF 57
    ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 IYGGTSSWGEWPMQVSLQVRLTAQRHLGCGSLGHQVWVTAHCFDGLPIQDVR--IY 58
OY 58 AGILQSFMEFYAGAY-QVEKVISHPNYSKTKNNDIAMKLRKPLTFNDLVKPVCLPNFG 116
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 59 SGILNLSITTDTPFSQIKETIIHQNVYSBGNHDIAMKLRQAPLVNVEFQKPLPSKG 118
OY 117 MLOPEQLCWSISGMKATEKTKTSEVLNAKAVLLLETORCNSRYVYDNLITPAMICAGFL 176
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 119 DSTIYTCWVITGWFSEKKEIQLKQVNIPLVTNECQKRY-QDKITQRYWCAQYK 177
OY 177 QGNVDSGDSGGLPVTSKNNIMWLLIGDTSWGSCKAKAVRPGVYGVNVVFTDWTYRQMR 235
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 178 EGGKDACRGDSGGLPVCKHNGMRLVGIITSGWEGCARREGQGVYTKVAEYMDWILEKTS 237
OY 236 ADG 238
    :||
DB 238 SDG 240

RESULT 4
US-08-681-151-3
; Sequence 3; Application US/08681151
```



```

Patent No.5869637
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Bandman, Olga
APPLICANT: Braxton, Scott Michael
APPLICANT: Goli, Surya
TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,151
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-007405
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 205011
US-08-681-151-3

Query Match 41.2%; Score 543; DB 2; Length 638;
Best Local Similarity 42.9%; Pred. No.4.9e-52;
Matches 103; Conservative 46; Mismatches 85; Indels 6; Gaps 4.

OY 1 RVGGSALPGMPQVSLHYQNV---HYCGGSITPEWIVTAAHQVEKPLNPMHWTA 57
DB 30 RVGGSSTNSLGGMPQVSLQVAVSONHMGSSITGRQVITLAAHCFD-GIPPDVWRL 448
OY 58 AGLTQSEFMFYAGY-QVEKVISHPNYSKTNNDIALMKLQKPLTFNDLVRPCLPN 116
DB 449 GGIILNLSEITNKTTPSSIKELIIHQKYMSEGSYDIALIKLTPLNTPEFQAPIC 508
OY 117 MMLQEPQLCWISGCGATEBKSGKTSEYLNAAKVLLETQRCNSRYVYDNLITPAM 176
DB 509 DNTIYITKCMWYGMGYTKERGETQNILOKATIPLVNECCQKKY-RDYVITK 567
OY 177 QGNVDSGCGDSCGLPYTSKNNNTIWLIGDTSWGSQCAKARPGVGVGMVFTD 236
DB 568 RGGIDACKGDSGGLVCKHSGRMLVGLTSMWEGCAKREQPGVYTKVAREYIDMI 627

```

```

: GENERAL INFORMATION:
: APPLICANT: O'Brien, Timothy J.
: APPLICANT: Tanimoto, Hirotsoshi
: TITLE OF INVENTION: TAG-15: An Extracellular Serine Protease Overexpressed in
: TITLE OF INVENTION: Breast and Ovarian Carcinomas
: FILE REFERENCE: D6064
: CURRENT APPLICATION NUMBER: US/09/027, 337B
: CURRENT FILING DATE: 1998-02-20
: NUMBER OF SEQ ID NOS: 13
: SEQ ID NO 3
: LENGTH: 256
: TYPE: PRT
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)
: OTHER INFORMATION: homologous to similar domain in TAG-15
US-09-027-337-3

Query Match          40.1%; Score 528.5; DB 2; Length 256;
Best Local Similarity 41.2%; Pred. No. 5,4e-51;
Matches 101; Conservative 41; Mismatches 92; Indels 11; Gaps 3;

QY      1 RIVGGSALPGAMPQVSLHYQNVHVGSGSIITPEMIVTAHGVKPEPLNPMHMTAFAGI 60
        |||||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||
DB      1 RIVGGWDTISLRPMQWVSLRYDGAHLCGSSLGSDWVLTAAHCFPEKNRYLSMRVAFAGA 60
QY      61 LRQSEMFYQAGYQVEKIVISHPNY-----DSKTRNDIALMKLQKPLTENDLVKPYCLPN 114
        :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      61 VAQASP-HQLQGLYQAVVYHGGILPFRDNRSENSNDIALVHLSPLPLEIYIQPYCLPA 119
QY      115 PCMMIQLPEQLCWISMGATEBEKGTSEVLNAKAVLLIEFQRCNSRYVDNLITPMACIG 174
        :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      120 AQAALVLDGKICVTGWTGNTQYVQQAQAVLQEARVPIISNDVCAGADFYGNQIRPKMECAG 179
QY      175 FLOGVNDSCQSGSGGLY----TSKNINIMWLIGDTSMSGSCAAKAPRGYGVNMFPTDWI 230
        :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      180 YPEGGIDACQSGSGPFCVEDSISRTPRMKLGIYSVSGTECALAQKPGVYTKVSDFERWTI 239
QY      231 YRQMR 235
        :  :  :
DB      240 PQAIR 244

RESULT 6
US-08-944-483-64
: Sequence 64, Application US/08944483
: Patent No. 6232456
: GENERAL INFORMATION:
: APPLICANT: COHEN, MAURICE
: APPLICANT: COLPITS, TRACEY L.
: APPLICANT: FRIEDMAN, PAULA N.
: APPLICANT: GRANADOS, EDWARD N.
: APPLICANT: KLAAS, MICHAEL R.
: APPLICANT: RUSSELL, JOHN C.
: APPLICANT: STEWART, KENT D.
: APPLICANT: STROUPE, STEVEN D.
: TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
: TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
: NUMBER OF SEQUENCES: 76
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Abbott Laboratories
: STREET: 100 Abbott Park Road
: CITY: Abbott Park
: STATE: IL
: COUNTRY: USA
: ZIP: 60064-3500
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0

```

RESULT 7
US-08-944-483-67
Sequence 67, Application US/08944483
Patent No. 6232456
GENERAL INFORMATION:
APPLICANT: COHEN, MAURICE
APPLICANT: COLPITS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GRANADOS, EDWARD N.
APPLICANT: KLASS, MICHAEL R.
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STEWART, KENT D.
APPLICANT: STROUPE, STEVEN D.
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
TITLE OF INVENTION: OF THE PROSTATE
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA

```

1 RESULT      8
2 US-08-200-900A-2
3 ; Sequence 2, Application US/08200900A
4 ; Patent No. 5665566
5 ;
6 ; GENERAL INFORMATION:
7 ;
8 ; APPLICANT:
9 ;
10 ; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
11 ;
12 ; NUMBER OF SEQUENCES: 38
13 ;
14 ; CORRESPONDENCE ADDRESS:
15 ;
16 ; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
17 ;
18 ; STREET: 87 Cambridgepark Drive
19 ;
20 ; CITY: Cambridge
21 ;
22 ; STATE: MA
23 ;
24 ; COUNTRY: USA
25 ;
26 ; ZIP: 02140
27 ;
28 ; COMPUTER READABLE FORM:
29 ;
30 ; MEDIUM TYPE: Floppy disk

```


Db 1 IVGSDSREGAMPVVALYFDQOVCASLSVSRDMLVSAHCVGRNMESKMAVGLH 60
QY 62 ROSFMFYAGAGYVE-----KVISHPNYDSKTKNDIALMKLOKPLTFNDLVKPVCLPBG 116
Db 61 MASNL---TSPOIETRLDQIYINPHYKRRKNNDIAMHLEMKRYNDYIPICLPEEN 117
QY 117 MMLQEPOLCMIWISGCGATEERKGTSEVLNAKAVLLIETORCNSRYVYDNLITPAMICAGF 176
Db 118 QVFPGRICISINGCALIYQGSTADVLQEDADVPLTSENKCOQOMPEYN-ITENMYCAGYE 176
QY 177 QGNVDSGCGDGGPLVTSKNNIMMLIGDTSWGSCKAKARPGVYGNVAFWTMI 230
Db 177 AGGVDSGCGDGGPLMKCEQNNRMLLAGVTSFGYQCALPNRPGVYARVPRTETMI 230

RESULT 11
US-09-008-271A-6
Sequence 6, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Hillman, Jennifer L.
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008.271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT13
CLONE: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6

Query Match 39.5%; Score 521; DB 4; Length 435;
Best Local Similarity 44.6%; Pred. No. 8e-50;
Matches 108; Conservative 36; Mismatches 84; Indels 14; Gaps 7;
QY 1 RIVGESALPGAMPQVSLVQNVHVCGSLITPEWITAAHCVEKPLNNPMHTAFAGI 60
||||| :||||| :||||| :||||| :|||

Db 202 RVVGESEASVDSMPQVSIQYDKQHYVCGSILDPHNVLTAAHCFRKH--TDVFNMKVRAGS 260
QY 61 LR-OSFMFYAGAGYVEKYV---SHPNYDSKTKNDIALMKLOKPLTFNDLVKPVCLPBG 116
Db 261 DKLGSF-----PSLAVAKIIIIIFENPMY---PKNDIALMKLOKPLTFSGYVAPICLPFD 313
QY 117 MMLQEPOLCMIWISGCGATEERKGTSEVLNAKAVLLIETORCNSRYVYDNLITPAMICAGF 175
Db 314 EELTPATPLMITIGWFTQNGGKMSDILLQASVYIDSTRCADAVAGGEVTEKMKCAGI 373
QY 176 LOGNVDSGCGDGGPLVTSKNNIMMLIGDTSWGSCKAKARPGVYGNVAFWTMI 235
Db 374 PEGGVDTGCGDGGPLMTQSDQ-WHVGVISWVGCGGPGVYTKVSAYLMIYVNWK 432
QY 236 AD 237
Db 433 AE 434

RESULT 12
US-08-508-448C-25
Sequence 25, Application US/08508448C
Patent No. 5804410
GENERAL INFORMATION:
APPLICANT: Kazuyoshi YAMAKA et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508.448C
FILING DATE: July 28, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-508-448C-25

Query Match 39.3%; Score 518.5; DB 1; Length 418;
Best Local Similarity 43.2%; Pred. No. 1.4e-49;
Matches 101; Conservative 34; Mismatches 94; Indels 5; Gaps 3;
QY 1 RIVGESALPGAMPQVSLVQNVHVCGSLITPEWITAAHCVEKPLNNPMHTAFAGI 60
||||| :||||| :||||| :||||| :|||
Db 186 RILGSTEAEBSWPMQVSLRLNNAHHCGLINNMILITAAHCF-RSSNSNRDMITATSGI 244
QY 61 ROSFMFYAGAGYVEKYVISHPNYDSKTKNDIALMKLOKPLTFNDLVKPVCLPBGMLQ 120

Db 245 ---STFPLRRNRNLLHNHNSKATHEMDIALVRLNSVTFKDIHSVCLPATONIP 301
QY 121 PDLCHMSWGATTEEGKTSVYLNAAKVLLEFQRNSRYVDNLITPAMICAGFLOGNV 180
Db 302 PGSTAYVTGWAQEVAGHTVPELRQGVRIISNDVCNASHSYNGAILSGMLCAGVPOGGV 361
QY 181 DSCGDSGGLV-TSKNNIMWLLIGDTSMGSCAKAYRPGVGNVMTWIRQ 233
Db 362 DACGDSGGLVQEDSRRLMFTVIGVSMGDCGLPDKPGYTRVATYLDIMRQ 415

RESULT 13

US-08-508-448C-19
; Sequence 19, Application US/08508448C
; Patent No. 5804410
; GENERAL INFORMATION:
; APPLICANT: Kazuyoshi YAMAOKA et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
; TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508.448C
; FILING DATE: July 28, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-508-448C-19

Query Match 38.9%; Score 513.5; DB 1; Length 232;

Best Local Similarity 42.9%; Pred. No. 2.2e-49;

Matches 100; Conservative 34; Mismatches 94; Indels 5; Gaps 3;

QY 2 YGGESALGAMPWOYSLHVNQVHCGGSLITPEWIVTAACHCEKPLNNPMTAFAGIL 61
Db 1 IIGGTAEIGSNPMQVSLNNNAHCGSLINMMWLTAAHCF-KSNSNPRDIATSGI- 58
QY 62 ROSFMETGAGYOVEKVISHPNDSKTKNDIALMKLQKPLTFENDLVKPYCLPDPGMLAP 121
Db 59 --STFPLRRNRNLLHNHNSKATHEMDIALVRLNSVTFKDIHSVCLPATONIP 116
QY 122 EQLCHMSWGATTEEGKTSVYLNAAKVLLEFQRNSRYVDNLITPAMICAGFLOGNV 181
Db 117 GSTAYVTGWAQEVAGHTVPELRQGVRIISNDVCNASHSYNGAILSGMLCAGVPOGGV 176

QY 182 SCGDSGGLV-TSKNNIMWLLIGDTSMGSCAKAYRPGVGNVMTWIRQ 233
Db 177 ACGDSGGLVQEDSRRLMFTVIGVSMGDCGLPDKPGYTRVATYLDIMRQ 229

RESULT 14

US-08-944-483-65
; Sequence 65, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLAAS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944.483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456
US-08-944-483-65

Query Match 38.8%; Score 511.5; DB 4; Length 235;

Best Local Similarity 41.6%; Pred. No. 3.8e-49;

Matches 96; Conservative 46; Mismatches 86; Indels 3; Gaps 2;

QY 2 YGGESALGAMPWOYSLHVNQVHCGGSLITPEWIVTAACHCEKPLNNPMTAFAGIL 61
Db 1 YGGSNAKEGAMPVWVGLLYGGRLLCGASLVSSDWLVSAAHCYVGRNLEPSKWTAILGLH 60
QY 62 ROSFMETGAGYOVEKVISHPNDSKTKNDIALMKLQKPLTFENDLVKPYCLPDPGML 119
Db 61 MKNLITSPQTVPLIDEIYINPYNRRKNDITAMHLEKVVYTYIOFTCLPEENOVF 120
QY 120 QPDLCHMSWGATTEEGKTSVYLNAAKVLLEFQRNSRYVDNLITPAMICAGFLOGNV 179

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:26:15 : Search time 164.81 Seconds
(without alignments)
320.367 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492

Perfect score: 1319
Sequence: 1 RIVGSGSALPGAMPQVSLH.....VCGNVWFYFDWIRQMRAD 238

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2803329 seqs, 221847457 residues

Total number of hits satisfying chosen parameters: 2803329

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/2/paa/PCRTUS.COMB.pep:*
- 2: /cgn2_6/ptodata/2/paa/US06.COMB.pep:*
- 3: /cgn2_6/ptodata/2/paa/US07.COMB.pep:*
- 4: /cgn2_6/ptodata/2/paa/US080.COMB.pep:*
- 5: /cgn2_6/ptodata/2/paa/US081.COMB.pep:*
- 6: /cgn2_6/ptodata/2/paa/US082.COMB.pep:*
- 7: /cgn2_6/ptodata/2/paa/US083.COMB.pep:*
- 8: /cgn2_6/ptodata/2/paa/US084.COMB.pep:*
- 9: /cgn2_6/ptodata/2/paa/US085.COMB.pep:*
- 10: /cgn2_6/ptodata/2/paa/US086.COMB.pep:*
- 11: /cgn2_6/ptodata/2/paa/US087.COMB.pep:*
- 12: /cgn2_6/ptodata/2/paa/US088.COMB.pep:*
- 13: /cgn2_6/ptodata/2/paa/US089.COMB.pep:*
- 14: /cgn2_6/ptodata/2/paa/US090.COMB.pep:*
- 15: /cgn2_6/ptodata/2/paa/US091.COMB.pep:*
- 16: /cgn2_6/ptodata/2/paa/US092.COMB.pep:*
- 17: /cgn2_6/ptodata/2/paa/US093.COMB.pep:*
- 18: /cgn2_6/ptodata/2/paa/US094.COMB.pep:*
- 19: /cgn2_6/ptodata/2/paa/US095.COMB.pep:*
- 20: /cgn2_6/ptodata/2/paa/US096.COMB.pep:*
- 21: /cgn2_6/ptodata/2/paa/US097.COMB.pep:*
- 22: /cgn2_6/ptodata/2/paa/US098.COMB.pep:*
- 23: /cgn2_6/ptodata/2/paa/US099.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1319	100.0	393	21	US-09-759-143-934
2	1319	100.0	393	21	US-09-780-669-934
3	1319	100.0	393	22	US-09-822-827-934
4	1319	100.0	393	22	US-09-852-911-934
5	1319	100.0	491	17	US-09-323-597-2
6	1319	100.0	492	1	PCT-US99-14622-2
7	1319	100.0	492	13	US-08-943-570-21
8	1319	100.0	492	17	US-08-323-597B-2
9	1319	100.0	492	20	US-09-615-285-2
10	1319	100.0	492	20	US-09-691-840-2

Result No.	Score	Query Match	Length	ID	Description
11	1319	100.0	492	21	US-09-759-143-932
12	1319	100.0	492	21	US-09-780-669-932
13	1319	100.0	492	22	US-09-822-827-932
14	1319	100.0	492	22	US-09-852-911-932
15	1319	100.0	492	18	US-09-408-628A-6
16	1314	99.6	384	23	US-60-239-841-23
17	1310	99.3	283	18	US-09-478-957-1
18	1303	98.8	250	1	PCT-US00-06111-8
19	1303	98.8	266	23	US-60-207-315-390
20	1303	98.8	266	23	US-60-230-435-1859
21	1303	98.8	492	1	PCT-US99-19655-15
22	1303	98.8	492	15	US-09-164-159-5
23	1303	98.8	492	17	US-09-323-597-4
24	1303	98.8	492	17	US-09-323-597B-4
25	1303	98.8	492	18	US-09-408-628B-5
26	1303	98.8	492	20	US-09-615-285-4
27	1303	98.8	492	20	US-09-679-426-895
28	1303	98.8	492	20	US-09-685-166-895
29	1303	98.8	492	21	US-09-759-143-895
30	1303	98.8	492	21	US-09-776-191-70
31	1303	98.8	492	21	US-09-780-669-895
32	1303	98.8	492	22	US-09-822-827-895
33	1303	98.8	492	22	US-09-852-911-895
34	1303	98.8	492	22	US-09-879-792-14
35	1303	98.8	605	23	US-60-212-659-375
36	1303	98.8	672	23	US-60-212-659-373
37	1090	82.6	227	23	US-60-230-435-1829
38	866	65.7	159	1	PCT-US00-05612-24
39	866	65.7	159	19	US-09-518-046-24
40	866	65.7	159	20	US-09-650-371-24
41	655	49.7	453	17	US-09-380-138-69
42	655	49.7	453	20	US-09-633-300-2
43	655	49.7	453	21	US-09-776-191-70
44	655	49.7	454	1	PCT-US00-05612-2
45	655	49.7	454	19	US-09-518-046-2

ALIGNMENTS

RESULT 1

US-09-759-143-934

Sequence 934, Application US/09759143

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun

APPLICANT: Dillon, Davin C.

APPLICANT: Mitcham, Jennifer L.

APPLICANT: Harlocker, Susan L.

APPLICANT: Jiang, Yuqi

APPLICANT: Henderson, Robert A.

APPLICANT: Kados, Michael D.

APPLICANT: Fanger, Gary R.

APPLICANT: Retter, Marc W.

APPLICANT: Stolk, John A.

APPLICANT: Day, Craig H.

APPLICANT: Vedvick, Thomas S.

APPLICANT: Carter, Darrick

APPLICANT: Li, Samuel

APPLICANT: Wang, Aljun

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Hepler, William

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

FILE REFERENCE: 210121.427C23

CURRENT APPLICATION NUMBER: US/09/759,143

CURRENT FILING DATE: 2001-01-12

NUMBER OF SEQ ID NOS: 934

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 934

LENGTH: 393

TYPE: PRT

ORGANISM: Homo sapiens

US-09-759-143-934

Query Match 100.0%; Score 1319; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 60
DB 156 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 215
OY 61 LROSFMEYAGAYOEKVIYSHPNYDSKTKNNNDIALMKLQKPLTFNDLVKPVCLPMPGMLQ 120
DB 216 LROSFMEYAGAYOEKVIYSHPNYDSKTKNNNDIALMKLQKPLTFNDLVKPVCLPMPGMLQ 275
OY 121 PEOLCWSGKATEEKGKTSSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLOGNV 180
DB 276 PEOLCWSGKATEEKGKTSSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLOGNV 335
OY 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKARPGVGNVVFDTWIRQMRADG 238
DB 336 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKARPGVGNVVFDTWIRQMRADG 393

RESULT 2

US-09-780-669-934
Sequence 934, Application US/09780669
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stoik, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C24
CURRENT APPLICATION NUMBER: US/09/780.669
CURRENT FILING DATE: 2001-02-09
NUMBER OF SEQ ID NOS: 943
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-780-669-934

Query Match 100.0%; Score 1319; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 60
DB 156 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 215
OY 61 LROSFMEYAGAYOEKVIYSHPNYDSKTKNNNDIALMKLQKPLTFNDLVKPVCLPMPGMLQ 120
DB 216 LROSFMEYAGAYOEKVIYSHPNYDSKTKNNNDIALMKLQKPLTFNDLVKPVCLPMPGMLQ 275

OY 121 PEOLCWSGKATEEKGKTSSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLOGNV 180
DB 276 PEOLCWSGKATEEKGKTSSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLOGNV 335
OY 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKARPGVGNVVFDTWIRQMRADG 238
DB 336 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKARPGVGNVVFDTWIRQMRADG 393

RESULT 3

US-09-822-827-934
Sequence 934, Application US/09822827
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stoik, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.534C1
CURRENT APPLICATION NUMBER: US/09/822.827
CURRENT FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-822-827-934

Query Match 100.0%; Score 1319; DB 22; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 60
DB 156 RIVGESALPGAMPQVSLHVQNVHVCSSITTPEMITAAHCVEKPLNNPMHTAFAGI 215
OY 61 LROSFMEYAGAYOEKVIYSHPNYDSKTKNNNDIALMKLQKPLTFNDLVKPVCLPMPGMLQ 120
DB 216 LROSFMEYAGAYOEKVIYSHPNYDSKTKNNNDIALMKLQKPLTFNDLVKPVCLPMPGMLQ 275
OY 121 PEOLCWSGKATEEKGKTSSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLOGNV 180
DB 276 PEOLCWSGKATEEKGKTSSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLOGNV 335
OY 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKARPGVGNVVFDTWIRQMRADG 238
DB 336 DSCGDSGGLVTSKNNIMWLIGDTSWGSCKAKARPGVGNVVFDTWIRQMRADG 393

RESULT 4

US-09-852-911-934
Sequence 934, Application US/09852911
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqi
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stoik, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C25
; CURRENT APPLICATION NUMBER: US/09/852,911
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 947
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-852-911-934
```

```
Query Match      100.0%; Score 1319; DB 22; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEVERKPLNNPMTAFAGI 60
DB 156 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEVERKPLNNPMTAFAGI 215
QY 61 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMLKQKPLTFNDLVKPYCLPFGMLQ 120
DB 216 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMLKQKPLTFNDLVKPYCLPFGMLQ 275
QY 121 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIQRCSRYVDNLITPAMICAGFLOGNV 180
DB 276 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIQRCSRYVDNLITPAMICAGFLOGNV 335
QY 181 DSCQGDGSGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGNVFTDWIYRQMRADG 238
DB 336 DSCQGDGSGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGNVFTDWIYRQMRADG 393
```

```
RESULT 5
US-09-323-597-2
; Sequence 2, Application US/09323597
```

```
; GENERAL INFORMATION:
; APPLICANT: Afari, Daniel E
; APPLICANT: Hubert, Rene S
; APPLICANT: Leong, Kahen
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Saifran, Douglas C.
; TITLE OF INVENTION: NOVEL TUMOR ANTIGEN USEFUL IN DIAGNOSIS AND THERAPY OF
; TITLE OF INVENTION: PROSTATE AND COLON CANCER
; FILE REFERENCE: 1703-007.US1
; CURRENT APPLICATION NUMBER: US/09/323,597
; CURRENT FILING DATE: 1999-06-01
; EARLIER APPLICATION NUMBER: 60/087,598
; EARLIER FILING DATE: 1998-06-01
; EARLIER APPLICATION NUMBER: 60/091,474
; EARLIER FILING DATE: 1998-06-29
; EARLIER APPLICATION NUMBER: 60/129,521
; EARLIER FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-323-597-2
```

```
Query Match      100.0%; Score 1319; DB 17; Length 491;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEVERKPLNNPMTAFAGI 60
DB 254 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEVERKPLNNPMTAFAGI 313
QY 61 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMLKQKPLTFNDLVKPYCLPFGMLQ 120
```

```
DB 314 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMLKQKPLTFNDLVKPYCLPFGMLQ 373
QY 121 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIQRCSRYVDNLITPAMICAGFLOGNV 180
DB 374 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIQRCSRYVDNLITPAMICAGFLOGNV 433
QY 181 DSCQGDGSGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGNVFTDWIYRQMRADG 238
DB 434 DSCQGDGSGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGNVFTDWIYRQMRADG 491
```

```
RESULT 6
PCT-US99-14622-2
; Sequence 2, Application PCT/US9914622
```

```
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: PCT/US99/14622
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US99-14622-2
```

```
Query Match      100.0%; Score 1319; DB 1; Length 492;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEVERKPLNNPMTAFAGI 60
DB 255 RIVGGSALPGAMPWOVSLHVONVHVGGSITTPETWITVAHCEVERKPLNNPMTAFAGI 314
QY 61 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMLKQKPLTFNDLVKPYCLPFGMLQ 120
DB 315 LROSEMFYAGYOVEKVISHPNDTSKTKNDIALMLKQKPLTFNDLVKPYCLPFGMLQ 374
QY 121 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIQRCSRYVDNLITPAMICAGFLOGNV 180
DB 375 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIQRCSRYVDNLITPAMICAGFLOGNV 434
QY 181 DSCQGDGSGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGNVFTDWIYRQMRADG 238
DB 435 DSCQGDGSGPLVTSKNNIMWLIGDTSWGSCKAKAYRPGVYGNVFTDWIYRQMRADG 492
```

```
RESULT 7
US-08-943-570-21
; Sequence 21, Application US/08943570
```

```
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASSEL, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STEWART, STEVEN D.
; TITLE OF INVENTION: NOVEL SPRING PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943.570
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6182, US, 01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
;
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: None
;
; US-08-943-570-21
;
Query Match      100.0%; Score 1319; DB 13; Length 492;
Best Local Similarity 100.0%; Pred. No. 6,6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHYQNVHVCSSITTPMIVTAACHCEKPLNNPWHMTAFAGI 60
DB 255 RIVGESALPGAMPQVSLHYQNVHVCSSITTPMIVTAACHCEKPLNNPWHMTAFAGI 314
QY 61 LROSFMFYGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNPGMMLQ 120
DB 315 LROSFMFYGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNPGMMLQ 374
QY 121 PEQLCWISGKATEEKGKTSSEVLNAKAVLLIETQRCNSRYVYDNLITPAMICAGFLQGNV 180
DB 375 PEQLCWISGKATEEKGKTSSEVLNAKAVLLIETQRCNSRYVYDNLITPAMICAGFLQGNV 434
QY 181 DSCGDSGGLPVTSKNNITWMLIGDTSWGSCKAKARPGVYGVNWVFTDWTIROMRAGD 238
DB 435 DSCGDSGGLPVTSKNNITWMLIGDTSWGSCKAKARPGVYGVNWVFTDWTIROMRAGD 492

RESULT 8
US-09-323-597B-2
; Sequence 2, Application US/09323597B
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; TITLE OF INVENTION: NOVEL TUMOR ANTIGEN USEFUL IN DIAGNOSIS
; TITLE OF INVENTION: AND THERAPY OF PROSTATE AND COLON CANCER
; FILE REFERENCE: 129, 80US01
; CURRENT APPLICATION NUMBER: US/09/323,597B
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,598
; PRIOR FILING DATE: 1998-06-01
```

```

; PRIOR APPLICATION NUMBER: 60/091,474
; PRIOR FILING DATE: 1998-06-29
; PRIOR APPLICATION NUMBER: 60/129,521
; PRIOR FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-09-323-597B-2
;
Query Match      100.0%; Score 1319; DB 17; Length 492;
Best Local Similarity 100.0%; Pred. No. 6,6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHYQNVHVCSSITTPMIVTAACHCEKPLNNPWHMTAFAGI 60
DB 255 RIVGESALPGAMPQVSLHYQNVHVCSSITTPMIVTAACHCEKPLNNPWHMTAFAGI 314
QY 61 LROSFMFYGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNPGMMLQ 120
DB 315 LROSFMFYGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNPGMMLQ 374
QY 121 PEQLCWISGKATEEKGKTSSEVLNAKAVLLIETQRCNSRYVYDNLITPAMICAGFLQGNV 180
DB 375 PEQLCWISGKATEEKGKTSSEVLNAKAVLLIETQRCNSRYVYDNLITPAMICAGFLQGNV 434
QY 181 DSCGDSGGLPVTSKNNITWMLIGDTSWGSCKAKARPGVYGVNWVFTDWTIROMRAGD 238
DB 435 DSCGDSGGLPVTSKNNITWMLIGDTSWGSCKAKARPGVYGVNWVFTDWTIROMRAGD 492

RESULT 9
US-09-615-285-2
; Sequence 2, Application US/09615285
; GENERAL INFORMATION:
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Stephen C. Mitchell
; APPLICANT: Aya Jakobovits
; APPLICANT: Mary Farris
; APPLICANT: Igor Vivanco
; TITLE OF INVENTION: NOVEL TUMOR ANTIGEN USEFUL IN DIAGNOSIS
; TITLE OF INVENTION: AND THERAPY OF PROSTATE AND COLON CANCER
; FILE REFERENCE: 129, 90US11
; CURRENT APPLICATION NUMBER: US/09/615,285
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 09/323,597
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo Sapiens
;
US-09-615-285-2
;
Query Match      100.0%; Score 1319; DB 20; Length 492;
Best Local Similarity 100.0%; Pred. No. 6,6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHYQNVHVCSSITTPMIVTAACHCEKPLNNPWHMTAFAGI 60
DB 255 RIVGESALPGAMPQVSLHYQNVHVCSSITTPMIVTAACHCEKPLNNPWHMTAFAGI 314
QY 61 LROSFMFYGAGYQVEKVISHPNYDSKTKNNDIALMKLQKPLTFNDLVKPVCLPNPGMMLQ 120
```

Db 315 LROSFMEYAGYOVEREYISHPNYDSKTKNNDIALMKLQKPLTFENDLYKPVCLPNCMMLQ 374
QY 121 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIETORCNSRYVYNLITTPAMICAGFLOGNV 180
Db 375 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIETORCNSRYVYNLITTPAMICAGFLOGNV 434
QY 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSACAKAYRPGVGNMVFETDWTYROMRADG 238
Db 435 DSCGDSGGLVTSKNNIMWLIGDTSWGSACAKAYRPGVGNMVFETDWTYROMRADG 492

RESULT 10
US-09-691-840-2
Sequence 932, Application US/09691840
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/691,840
CURRENT FILING DATE: 2000-10-18
PRIOR APPLICATION NUMBER: US/09/342,749
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: US 60/091,044
PRIOR FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-691-840-2

Query Match 100.0%; Score 1319; DB 20; Length 492;
Best Local Similarity 100.0%; Pred. No. 6,6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGGSALPGAMPQVSLHVQVHVGSGSIITPEWITVAHCVEKPLNPMHTAFAGI 60
Db 255 RIVGGSALPGAMPQVSLHVQVHVGSGSIITPEWITVAHCVEKPLNPMHTAFAGI 314
QY 61 LROSFMEYAGYOVEREYISHPNYDSKTKNNDIALMKLQKPLTFENDLYKPVCLPNCMMLQ 120
Db 315 LROSFMEYAGYOVEREYISHPNYDSKTKNNDIALMKLQKPLTFENDLYKPVCLPNCMMLQ 374
QY 121 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIETORCNSRYVYNLITTPAMICAGFLOGNV 180
Db 375 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIETORCNSRYVYNLITTPAMICAGFLOGNV 434
QY 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSACAKAYRPGVGNMVFETDWTYROMRADG 238
Db 435 DSCGDSGGLVTSKNNIMWLIGDTSWGSACAKAYRPGVGNMVFETDWTYROMRADG 492

RESULT 11
US-09-759-143-932
Sequence 932, Application US/09759143
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolck, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.

APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C23
CURRENT APPLICATION NUMBER: US/09/759,143
CURRENT FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-759-143-932

Query Match 100.0%; Score 1319; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 6,6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGGSALPGAMPQVSLHVQVHVGSGSIITPEWITVAHCVEKPLNPMHTAFAGI 60
Db 255 RIVGGSALPGAMPQVSLHVQVHVGSGSIITPEWITVAHCVEKPLNPMHTAFAGI 314
QY 61 LROSFMEYAGYOVEREYISHPNYDSKTKNNDIALMKLQKPLTFENDLYKPVCLPNCMMLQ 120
Db 315 LROSFMEYAGYOVEREYISHPNYDSKTKNNDIALMKLQKPLTFENDLYKPVCLPNCMMLQ 374
QY 121 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIETORCNSRYVYNLITTPAMICAGFLOGNV 180
Db 375 PEOLCWTISGKATEEKGKTSSEVLNAKVLLEIETORCNSRYVYNLITTPAMICAGFLOGNV 434
QY 181 DSCGDSGGLVTSKNNIMWLIGDTSWGSACAKAYRPGVGNMVFETDWTYROMRADG 238
Db 435 DSCGDSGGLVTSKNNIMWLIGDTSWGSACAKAYRPGVGNMVFETDWTYROMRADG 492

RESULT 12
US-09-780-669-932
Sequence 932, Application US/09780669
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolck, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C24
CURRENT APPLICATION NUMBER: US/09/780,669
CURRENT FILING DATE: 2001-02-09
NUMBER OF SEQ ID NOS: 943
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-780-669-932

```

```

Query Match      100.0%; Score 1319; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKPLNNPWHMTAFAGI 60
    |||||||
Db 255 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKPLNNPWHMTAFAGI 314
    |||||||
Oy 61 LRQSFMYGAGYQVEKVISHPYDSTKNDIALMKLQPLTFNDLVKPVCLPFGGMILQ 120
    |||||||
Db 315 LRQSFMYGAGYQVEKVISHPYDSTKNDIALMKLQPLTFNDLVKPVCLPFGGMILQ 374
    |||||||
Oy 121 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
    |||||||
Db 375 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 434
    |||||||
Oy 181 DSCQDSSGGLVTSKNNIMWLIGDTSWSSGCAKAYRPGYGVNVFTDWTYROMRADG 238
    |||||||
Db 435 DSCQDSSGGLVTSKNNIMWLIGDTSWSSGCAKAYRPGYGVNVFTDWTYROMRADG 492
    |||||||

```

RESULT 13

```

US-09-822-827-932
; Sequence 932, Application US/09822827
; GENERAL INFORMATION:

```

```

; APPLICANT: Xu, Jiaqichun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-822-827-932

```

```

Query Match      100.0%; Score 1319; DB 22; Length 492;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKPLNNPWHMTAFAGI 60
    |||||||
Db 255 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKPLNNPWHMTAFAGI 314
    |||||||
Oy 61 LRQSFMYGAGYQVEKVISHPYDSTKNDIALMKLQPLTFNDLVKPVCLPFGGMILQ 120
    |||||||
Db 315 LRQSFMYGAGYQVEKVISHPYDSTKNDIALMKLQPLTFNDLVKPVCLPFGGMILQ 374
    |||||||
Oy 121 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
    |||||||
Db 375 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 434
    |||||||
Oy 181 DSCQDSSGGLVTSKNNIMWLIGDTSWSSGCAKAYRPGYGVNVFTDWTYROMRADG 238
    |||||||
Db 435 DSCQDSSGGLVTSKNNIMWLIGDTSWSSGCAKAYRPGYGVNVFTDWTYROMRADG 492
    |||||||

```

RESULT 14

```

US-09-852-911-932
; Sequence 932, Application US/09852911
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiaqichun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.

```

```

; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darlick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C25
; CURRENT APPLICATION NUMBER: US/09/852,911
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 947
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 932
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-852-911-932

```

```

Query Match      100.0%; Score 1319; DB 22; Length 492;
Best Local Similarity 100.0%; Pred. No. 6.6e-131;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKPLNNPWHMTAFAGI 60
    |||||||
Db 255 RIVGESALPGAMPQVSLHVONVHVCSSITTPETVITAACHVEKPLNNPWHMTAFAGI 314
    |||||||
Oy 61 LRQSFMYGAGYQVEKVISHPYDSTKNDIALMKLQPLTFNDLVKPVCLPFGGMILQ 120
    |||||||
Db 315 LRQSFMYGAGYQVEKVISHPYDSTKNDIALMKLQPLTFNDLVKPVCLPFGGMILQ 374
    |||||||
Oy 121 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 180
    |||||||
Db 375 PEOLCWMISGWTGTEKGTSEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNV 434
    |||||||
Oy 181 DSCQDSSGGLVTSKNNIMWLIGDTSWSSGCAKAYRPGYGVNVFTDWTYROMRADG 238
    |||||||
Db 435 DSCQDSSGGLVTSKNNIMWLIGDTSWSSGCAKAYRPGYGVNVFTDWTYROMRADG 492
    |||||||

```

RESULT 15

```

US-09-408-628A-6
; Sequence 6, Application US/09408628A
; GENERAL INFORMATION:
; APPLICANT: Macbeth, Kyle J.
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NUCLEIC ACIDS AND POLYPEPTIDES USEFUL IN
; FILE REFERENCE: 0734-135001
; CURRENT APPLICATION NUMBER: US/09/408,628A
; CURRENT FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: US 09/164,159
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: US 09/163,759
; PRIOR FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-09-408-628A-6

Query Match 99.8%; Score 1316; DB 18; Length 492;
Best Local Similarity 99.6%; Pred. No. 1.4e-130;
Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	RIVGESALPGAMPQVSLHVQNVHVCGSIIIPENIVTAHCEKPLNNPWHMTAFAGI	60
DB	255	RIVGESALPGAMPQVSLHVQNVHVCGSIIIPENIVTAHCEKPLNNPWHMTAFAGI	314
QY	61	LROSEMFYAGYQVEKVISHPNYSKTKNNDIALLMKLQKPLTFNDLVKPYCLPMPGMLQ	120
DB	315	LROSEMFYAGYQVEKVISHPNYSKTKNNDIALLMKLQKPLTFNDLVKPYCLPMPGMLQ	374
QY	121	PROLCWISGNGATEEKGKTSVLNAKVLIIETQRCNSRYVDNLITPAMICAGFLQGNV	180
DB	375	PROLCWISGNGATEEKGKTSVLNAKVLIIETQRCNSRYVDNLITPAMICAGFLQGNV	434
QY	181	DSQGDGSGPLVTSKNNIMWLLIGDTSMSGSCAKAYRPGYGNMVFPTDNIYROMRADG	238
DB	435	DSQGDGSGPLVTSKNNIMWLLIGDTSMSGSCAKAYRPGYGNMVFPTDNIYROMRADG	492

Search completed: September 26, 2001, 09:26:16
Job time: 311 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:23:22 : Search time 27.32 Seconds
(without alignments)
370.445 Million cell updates/sec

Title: US-09-615-285-2_COPY_255_492
Perfect score: 1319
Sequence: 1 RIVGESALPGAMPQVSLH.....VYGNVVFETDVIYRQMRADG 238

Scoring table:
BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 277479 seqs, 42523304 residues

Total number of hits satisfying chosen parameters: 277479

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/1/paa/PCF_NEW_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1319	100.0	260	US-09-760-475-2355	Sequence 2355, Ap
2	1319	100.0	393	US-09-895-793-934	Sequence 934, App
3	1319	100.0	393	US-09-895-814-934	Sequence 934, App
4	1319	100.0	492	US-09-895-793-932	Sequence 932, App
5	1319	100.0	492	US-09-895-814-932	Sequence 932, App
6	1303	98.8	492	US-09-895-793-895	Sequence 895, App
7	1303	98.8	492	US-09-895-814-895	Sequence 895, App
8	1263	95.8	229	US-09-715-427-53	Sequence 53, App
9	894	67.8	163	US-09-760-475-3414	Sequence 3414, Ap
10	645.5	48.9	248	US-09-925-301-1017	Sequence 1017, Ap
11	645.5	48.9	394	PCF-US01-18569-3126	Sequence 3126, Ap
12	561	42.1	421	US-09-715-427-15	Sequence 15, Appl
13	529.5	40.1	343	US-09-875-195-922	Sequence 922, App
14	523.5	39.7	255	US-09-410-362B-6	Sequence 6, Appl
15	521	39.5	435	PCF-US01-18568-2	Sequence 2, Appl
16	521	39.5	435	PCF-US01-18568A-2	Sequence 2, Appl
17	511.5	38.8	235	US-09-410-362B-5	Sequence 5, Appl
18	506	38.6	521	US-09-764-898-216	Sequence 216, App
19	506	38.4	292	PCF-US01-18568-9	Sequence 9, Appl
20	506	38.4	292	PCF-US01-18568A-9	Sequence 9, Appl
21	502.5	38.1	812	US-09-457-018-81	Sequence 81, Appl
22	502	38.1	236	US-09-715-427-44	Sequence 44, Appl
23	497	37.7	855	US-09-410-362B-2	Sequence 2, Appl
24	477	36.2	241	US-09-410-362B-3	Sequence 3, Appl
25	466	35.3	932	PCF-US01-08631-51801	Sequence 51801, A
26	466	35.3	916	US-09-758-449-1401	Sequence 1401, Ap
27	465.5	35.3	293	US-09-764-898-213	Sequence 213, Ap

28	465.5	35.3	449	1	PCF-US01-08631-58151	Sequence 58151, A
29	458	34.7	413	5	US-09-715-427-11	Sequence 11, Appl
30	457.5	34.7	305	5	US-09-885-441-57	Sequence 57, Appl
31	457.5	34.7	334	5	US-09-885-441-56	Sequence 56, Appl
32	455.5	34.5	244	5	US-09-715-427-51	Sequence 51, Appl
33	455.5	34.5	314	5	US-09-885-441-14	Sequence 14, Appl
34	455.5	34.5	314	5	US-09-664-610B-257	Sequence 257, App
35	455.5	34.5	317	5	US-09-882-636-23	Sequence 23, Appl
36	455.5	34.5	317	5	US-09-664-610B-263	Sequence 263, App
37	455.5	34.5	325	5	US-09-764-898-278	Sequence 278, App
38	455.5	34.5	325	5	US-09-764-902-1211	Sequence 1211, Ap
39	455.5	34.5	325	5	US-09-908-711-114	Sequence 114, App
40	455.5	34.5	327	5	US-09-787-844-2	Sequence 2, Appl
41	455.5	34.5	351	5	US-09-764-898-207	Sequence 207, App
42	454.5	34.5	322	5	US-09-764-898-279	Sequence 279, App
43	454.5	34.5	322	5	US-09-764-902-1147	Sequence 1147, Ap
44	452.5	34.3	230	5	US-09-760-475-2331	Sequence 2331, Ap
45	451	34.2	260	5	US-09-885-441-58	Sequence 58, Appl

ALIGNMENTS

RESULT 1
US-09-760-475-2355
Sequence 2355, Application US/09760475
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT249
CURRENT APPLICATION NUMBER: US/09/760,475
CURRENT FILING DATE: 2001-01-16
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 4122
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2355
LENGTH: 260
TYPE: PRT
ORGANISM: Homo sapiens
US-09-760-475-2355

Query Match 100.0%; Score 1319; DB 5; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.3e+120;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RIVGESALPGAMPQVSLHVNVAHVGGSITPEMITVTAHCEVERPLNNPMTAFAGI 60
DB 23 RIVGESALPGAMPQVSLHVNVAHVGGSITPEMITVTAHCEVERPLNNPMTAFAGI 82
QY 61 LRQSFMYGAGYVEKVISHPNDSTKKNDDIALMLQKPLTFNDLVKPYCLPQGMILQ 120
DB 83 LRQSFMYGAGYVEKVISHPNDSTKKNDDIALMLQKPLTFNDLVKPYCLPQGMILQ 142
QY 121 PEOLCMTSGGATEEGKTSSEVLNAKVLLETFQCNRSRYVNYLITPAMICAGFLQGNV 180
DB 143 PEOLCMTSGGATEEGKTSSEVLNAKVLLETFQCNRSRYVNYLITPAMICAGFLQGNV 202
QY 181 DSCQDSGGLVTSKNNIMWLIGDTSWGSQCAKARPGYGVNMFETDIYRQMRADG 238
DB 203 DSCQDSGGLVTSKNNIMWLIGDTSWGSQCAKARPGYGVNMFETDIYRQMRADG 260

RESULT 2
US-09-895-793-934
Sequence 934, Application US/09895793
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Kalos, Michael D.

```

: APPLICANT: Retter, Marc W.
: APPLICANT: Stolk, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darlick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aljun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hural, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
: APPLICANT: Fanger, Gary R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.534C2
: CURRENT APPLICATION NUMBER: US/09/895,793
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 982
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 934
: LENGTH: 393
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-895-793-934
```

```

Query Match          100.0%; Score 1319; DB 5; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.2e-120;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RIVGESALPGAMPQVSLHVQNVHVCSSITTPKMTVTAACHVEKPLNNPWHMTAFAGI 60
Db 156 RIVGESALPGAMPQVSLHVQNVHVCSSITTPKMTVTAACHVEKPLNNPWHMTAFAGI 215

Oy 61 LRQSFMYGAGYQVEKVI SHPNYDSKTKNNDIALMLKLRPLTFNDLVKPCLPNPGMMLQ 120
Db 216 LRQSFMYGAGYQVEKVI SHPNYDSKTKNNDIALMLKLRPLTFNDLVKPCLPNPGMMLQ 275

Oy 121 PEQLCWMISGWTGTEKGTSEVLNAKAVLLIETQRCNSRYVDNLITPAMICAGELQGNV 180
Db 276 PEQLCWMISGWTGTEKGTSEVLNAKAVLLIETQRCNSRYVDNLITPAMICAGELQGNV 335

Oy 181 DSCQDSSGPLVTSKNNTMMLIGDTSMGSGCAKAYRPGYGVNVFTDWTIRQMRADG 238
Db 336 DSCQDSSGPLVTSKNNTMMLIGDTSMGSGCAKAYRPGYGVNVFTDWTIRQMRADG 393
```

```

RESULT 3
US-09-895-814-934
: Sequence 934, Application US/09895814
: GENERAL INFORMATION:
: APPLICANT: Xu, Jlangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jlang, Yuguu
: APPLICANT: Kalos, Michael D.
: APPLICANT: Retter, Marc W.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darlick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aljun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hural, John
: APPLICANT: McNeill, Patricia D.
```

```

: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
: APPLICANT: Fanger, Gary R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.427C26
: CURRENT APPLICATION NUMBER: US/09/895,814
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 990
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 934
: LENGTH: 393
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-895-814-934
```

```

Query Match          100.0%; Score 1319; DB 5; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.2e-120;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RIVGESALPGAMPQVSLHVQNVHVCSSITTPKMTVTAACHVEKPLNNPWHMTAFAGI 60
Db 156 RIVGESALPGAMPQVSLHVQNVHVCSSITTPKMTVTAACHVEKPLNNPWHMTAFAGI 215

Oy 61 LRQSFMYGAGYQVEKVI SHPNYDSKTKNNDIALMLKLRPLTFNDLVKPCLPNPGMMLQ 120
Db 216 LRQSFMYGAGYQVEKVI SHPNYDSKTKNNDIALMLKLRPLTFNDLVKPCLPNPGMMLQ 275

Oy 121 PEQLCWMISGWTGTEKGTSEVLNAKAVLLIETQRCNSRYVDNLITPAMICAGELQGNV 180
Db 276 PEQLCWMISGWTGTEKGTSEVLNAKAVLLIETQRCNSRYVDNLITPAMICAGELQGNV 335

Oy 181 DSCQDSSGPLVTSKNNTMMLIGDTSMGSGCAKAYRPGYGVNVFTDWTIRQMRADG 238
Db 336 DSCQDSSGPLVTSKNNTMMLIGDTSMGSGCAKAYRPGYGVNVFTDWTIRQMRADG 393
```

```

RESULT 4
US-09-895-793-932
: Sequence 932, Application US/09895793
: GENERAL INFORMATION:
: APPLICANT: Xu, Jlangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jlang, Yuguu
: APPLICANT: Kalos, Michael D.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stolk, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darlick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aljun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hural, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
: APPLICANT: Fanger, Gary R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.534C2
: CURRENT APPLICATION NUMBER: US/09/895,793
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 982
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 932
```

```

; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-932

```

Query Match	100.0%;	Score 1319;	DB 5;	Length 492;
Best Local Similarity	100.0%;	Pred. NO. 2.9e-120;		
Matches 238;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	RIVGESSLPAMPNOVSLHQNWHGGSSTIREPVLTAHCYERKLNPNHMTAFAGI	60
Db	255	RIVGESSLPAMPNOVSLHQNWHGGSSTIREPVLTAHCYERKLNPNHMTAFAGI	314
QY	61	LROSFMEYAGYQVEKYLSHPNVDSTKRNNDIALMLQKRLTENDLVKPYCLSPNGMILQ	120
Db	315	LROSFMEYAGYQVEKYLSHPNVDSTKRNNDIALMLQKRLTENDLVKPYCLSPNGMILQ	374
QY	121	PEOLCWLISGMATEKEKGTSEVLLAAAVLLIETORCNSRIVYDNLITPAMICAGFLOGNV	180
Db	375	PEOLCWLISGMATEKEKGTSEVLLAAAVLLIETORCNSRIVYDNLITPAMICAGFLOGNV	434
QY	181	DSOGDSDSGPLVTSKNNINIMWLIGTSMGSGCAKARYRGVGVNWVFDWLTIRQMRADG	238
Db	435	DSOGDSDSGPLVTSKNNINIMWLIGTSMGSGCAKARYRGVGVNWVFDWLTIRQMRADG	492

RESULT 5
US-09-895-814-932

```

APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuguu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Ranger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C26
CURRENT APPLICATION NUMBER: US/09/895,814
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 90
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-814-932

```

```

Query Match      100.0%; Score 1319; DB 5; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.9e-120;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY      1 RIVGESALPGAMPQVSLHVNQNHVCGSLTTPENIYTAHCVKEPLNNPHWTFAGI 60
|||||

```

Db	255	RIVGSESLPCGAMPRQVSLHVQNVHVGGSILITPEWVITAAHCVEKPLRPNNPHTATAGI	314
Qy	61	LRSMPFGAGYQVEKVTSHHPNYSKTKRNDIALMKLQPLTEPNDLYKPVCLPYPGMILO	122
Db	315	LRSMPFGAGYQVEKVTSHHPNYSKTKRNDIALMKLQPLTEPNDLYKPVCLPYPGMILO	374

Oy	121	PEOLCISMGWCAETBEKCKGTSVLNNAKVLLETPGRNSRYNDLITPAMICAELOGNV	180
		PEOLCISMGWCAETBEKCKGTSVLNNAKVLLETPGRNSRYNDLITPAMICAELOGNV	
		PEOLCISMGWCAETBEKCKGTSVLNNAKVLLETPGRNSRYNDLITPAMICAELOGNV	
Db	375	PEOLCISMGWCAETBEKCKGTSVLNNAKVLLETPGRNSRYNDLITPAMICAELOGNV	433
		PEOLCISMGWCAETBEKCKGTSVLNNAKVLLETPGRNSRYNDLITPAMICAELOGNV	
Oy	181	DSCGDSGGLPLVTSKNNITWMLIGDTSMGSGCAKAPREVYGVNMYEFDWITYRQRADG	238
		DSCGDSGGLPLVTSKNNITWMLIGDTSMGSGCAKAPREVYGVNMYEFDWITYRQRADG	
Db	435	DSCGDSGGLPLVTSKNNITWMLIGDTSMGSGCAKAPREVYGVNMYEFDWITYRQRADG	492
		DSCGDSGGLPLVTSKNNITWMLIGDTSMGSGCAKAPREVYGVNMYEFDWITYRQRADG	

RESULT 6
US-09-895-793-895
; Sequence 895, Application US/09895793

```

: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jiang, Yugu
: APPLICANT: Kalos, Michael D.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stolk, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darrick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aijun
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hurral, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
: APPLICANT: Fanger, Gary R.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.534C2
: CURRENT APPLICATION NUMBER: US/09/895,793
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 982
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 895
: LENGTH: 492
: TYPE: prf
: ORGANISM: Homo sapiens
: US-09-895-793-895

```

Query Match	98.8%;	Score 1303;	DB 5;	Length 492;
Best Local Similarity	98.3%;	Pred. No. 1e-118;		
Matches 234;	Conservative 3;	Mismatches 1;	Indels 0;	Gaps 0;

Qy	1	RIVGESSLPCAMPWQVSLHVNQVHGGSSITIPENITVAAHCEVPLNNPHMTAFAGI	
Db	255	RIVGESSLPCAMPWQVSLHVNQVHGGSSITIPENITVAAHCEVPLNNPHMTAFAGI	3
Qy	61	LROSMFPGAGYQVEKYISHPNTDSKTKNNDIAIMKLQKPLTFNDLVKVCCLPFGMLO	1
Db	315	LROSMFPGAGYQVQKVISHPNTDSKTKNNDIAIMKLQKPLTFNDLVKVCCLPFGMLO	3
Qy	121	PEOLCMISGMWATEEKEKTSSEVLAAVLLIEFQRNSRYVDNLITTPMIAQGLQVNG	1
Db	375	PEOLCMISGMWATEEKEKTSSEVLAAVLLIEFQRNSRYVDNLITTPMIAQGLQVNG	4
Qy	181	DSGGDSGGPLVTSKNNIMWLIGDTSMGSCCAATRRGYGVNMFETDVIYQMRADG	238

```

Db      435 DSCGDSGGPLVTSNNIMWLIIGDTSWGSCKAKAYRPGVGNVWFTDWTYROMKANG 492

RESULT 7
US-09-895-814-895
; Sequence 895, Application US/09895814
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Ketter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurst, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-895

Query Match      98.8%; Score 1303; DB 5; Length 492;
Best Local Similarity 98.3%; Pred. No. 1e-118;
Matches 234; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY      1 RIVGESALPGAMPWQVSLHVONVHVGSGSITTPEMVYTAACHVEKPLNNPWHMTAFAGI 60
Db      255 RIVGESALPGAMPWQVSLHVONVHVGSGSITTPEMVYTAACHVEKPLNNPWHMTAFAGI 314
OY      61 LRQSFMEYAGYQVEKVISHPNYDSKTRKNDIALMKLQKPLTFNDLVKPVCLPNEGMMLO 120
Db      315 LRQSFMEYAGYQVEKVISHPNYDSKTRKNDIALMKLQKPLTFNDLVKPVCLPNEGMMLO 374
OY      121 PEQLCWISGKATEEKGKTSVLAIAKYLLEIETQRCNSRYVDNLITPAMICAGFLQGNV 180
Db      375 PEQLCWISGKATEEKGKTSVLAIAKYLLEIETQRCNSRYVDNLITPAMICAGFLQGNV 434
OY      181 DSCGDSGGPLVTSNNIMWLIIGDTSWGSCKAKAYRPGVGNVWFTDWTYROMKANG 238
Db      435 DSCGDSGGPLVTSNNIMWLIIGDTSWGSCKAKAYRPGVGNVWFTDWTYROMKANG 492

RESULT 8
US-09-715-427-53
; Sequence 53, Application US/09715427
; GENERAL INFORMATION:
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Majumder, Kumud
; APPLICANT: Vermet, Corine
; APPLICANT: Herrmann, John L.

```

```

; APPLICANT: Burgess, Catherine
; APPLICANT: Fernandes, Elma
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Rastelli, Luca
; APPLICANT: Curagen Corporation
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-598
; CURRENT APPLICATION NUMBER: US/09/715,427
; CURRENT FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/165,986
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,839
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/195,637
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/197,080
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/232,677
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/181,347
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/194,195
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 53
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-715-427-53

Query Match      95.8%; Score 1263; DB 5; Length 229;
Best Local Similarity 99.1%; Pred. No. 3.1e-115;
Matches 227; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY      2 IVGESALPGAMPWQVSLHVONVHVGSGSITTPEMVYTAACHVEKPLNNPWHMTAFAGIL 61
Db      1 IVGESALPGAMPWQVSLHVONVHVGSGSITTPEMVYTAACHVEKPLNNPWHMTAFAGIL 60
OY      62 ROSFMEYAGYQVEKVISHPNYDSKTRKNDIALMKLQKPLTFNDLVKPVCLPNEGMMLOP 121
Db      61 ROSFMEYAGYQVEKVISHPNYDSKTRKNDIALMKLQKPLTFNDLVKPVCLPNEGMMLOP 120
OY      122 EQLCWISGKATEEKGKTSVLAIAKYLLEIETQRCNSRYVDNLITPAMICAGFLQGNV 181
Db      121 EQLCWISGKATEEKGKTSVLAIAKYLLEIETQRCNSRYVDNLITPAMICAGFLQGNV 180
OY      182 SCQDSDGGPLVTSNNIMWLIIGDTSWGSCKAKAYRPGVGNVWFTDWTY 230
Db      181 SCQDSDGGPLVTSNNIMWLIIGDTSWGSCKAKAYRPGVGNVWFTDWTY 229

RESULT 9
US-09-760-475-3414
; Sequence 3414, Application US/09760475
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT249
; CURRENT APPLICATION NUMBER: US/09/760,475
; CURRENT FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4122
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3414
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-760-475-3414

Query Match      67.8%; Score 894; DB 5; Length 163;

```



```
Db 179 RIVGALSDSKMPQVSLHFGTTHICGTLDAQWVLTAAHCFVTRKVEG--WKV 235
QY 57 FAGISQSFMEYAGYQVEKYVISHPNYDSKTKNNIDIALMKLOKPLTFNDLVPCVLPNG 116
Db 236 YAGTNSNLHQLEBA--STAEIINSNYDEEDDYDIALMKRLSKPLTSLAHIPACLPNHG 293
QY 117 MMLQPEQLCMTISGWGATEE-KGTSEVLNAKVLLEIETORNSRYVNDLITPAMICAGF 175
Db 294 QTFSLNETCMTWTFGKTRTETDCKTSPFLREVQVNLIDFKKCDYLVDSYLTTPRMKACAD 353
QY 176 LOGNVDSQSGSGGPLYSKNNIMWLIGDTSWGSCKAKAYRPGVGNMVFETDVIYRQMR 235
Db 354 LRGGDSQSGSGGPFVCEQNNRMVLAGVTSWGTGCGGRNKPVGVTYKTEVLPMTYSKME 413
QY 236 AD 237
Db 414 SE 415
```

```
RESULT 13
US-09-875-195-922
: Sequence 922, Application US/09875195
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA132N2
: CURRENT APPLICATION NUMBER: US/09/875,195
: PRIORITY FILING DATE: 2001-06-07
: Prior application data removed - consult PAM or file wrapper
: NUMBER OF SEQ ID NOS: 922
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 922
: LENGTH: 343
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (309)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-875-195-922
```

```
Query Match 40.1%; Score 529.5; DB 5; Length 343;
Best Local Similarity 41.2%; Pred. No. 1.1e-43;
Matches 101; Conservative 41; Mismatches 92; Indels 11; Gaps 3;

QY 1 RIVGESALPGAMPQVSLHVNHYVCGSITTPETVTAHCHVEKPLNNPMHTAFAGI 60
Db 88 RIVGGRDTSIGRMPQVSLRYDGAHLGGSLSGDMVLTAAHCFERRRNVLSRMKVFAG 147
QY 61 LRQSFMEYAGYQVEKYVISHPNY-----DSKTKNNIDIALMKLOKPLTFNDLVPCVLPN 114
Db 148 VAAQSP-HGLQGVQAVVYHGGYLPFRDPNSEBNSNDIALVHLSPLPLEYIOPVCLPA 206
QY 115 PGMALQPEQLCMTISGWGATEEKTSEVLNAKVLLEIETORNSRYVNDLITPAMICAG 174
Db 207 AGQALVDEKICTYVGMNTQYVGOAGVLAQARVPIISNDVCGADFGNIGKPKMFCAG 266
QY 175 FLQGVNDSQSGSGGPLY-----TSKNNIMWLIGDTSWGSCKAKAYRPGVGNMVFETDVI 230
Db 267 YPEGIDACQSGSGPFVCEQNNRMVLAGVTSWGTGCGGRNKPVGVTYKTEVLPMTYSKME 326
QY 231 YRQMR 235
Db 327 FOAIR 331
```

```
RESULT 14
US-09-410-362B-6
: Sequence 6, Application US/09410362B
: GENERAL INFORMATION:
: APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
```

```
: APPLICANT: Craik, Charles
: APPLICANT: Takeuchi, Toshiko
: TITLE OF INVENTION: MEMBRANE TYPE SERINE PROTEASE 1 (MT-SP10 AND USES THEREOF)
: FILE REFERENCE: 305T-9001-000S
: CURRENT APPLICATION NUMBER: US/09/410,362B
: PRIORITY FILING DATE: 1999-09-30
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 6
: LENGTH: 255
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Protein fragment/domain
US-09-410-362B-6
```

```
Query Match 39.7%; Score 523.5; DB 5; Length 255;
Best Local Similarity 41.0%; Pred. No. 3e-43;
Matches 100; Conservative 41; Mismatches 92; Indels 11; Gaps 3;

QY 2 IYGGESALPGAMPQVSLHVNHYVCGSITTPETVTAHCHVEKPLNNPMHTAFAGI 61
Db 1 IYGGGRDTSIGRMPQVSLRYDGAHLGGSLSGDMVLTAAHCFERRRNVLSRMKVFAG 60
QY 62 ROSFMEYAGYQVEKYVISHPNY-----DSKTKNNIDIALMKLOKPLTFNDLVPCVLPN 115
Db 61 AAQSP-HGLQGVQAVVYHGGYLPFRDPNSEBNSNDIALVHLSPLPLEYIOPVCLPA 119
QY 116 GMLQPEQLCMTISGWGATEEKTSEVLNAKVLLEIETORNSRYVNDLITPAMICAGF 175
Db 120 GOALVDEKICTYVGMNTQYVGOAGVLAQARVPIISNDVCGADFGNIGKPKMFCAGY 179
QY 176 LOGNVDSQSGSGGPLY-----TSKNNIMWLIGDTSWGSCKAKAYRPGVGNMVFETDVI 231
Db 180 PEGIDACQSGSGGPFVCEQNNRMVLAGVTSWGTGCGGRNKPVGVTYKTEVLPMTYSKME 239
QY 232 RQMR 235
Db 240 QAIR 243
```

```
RESULT 15
PCT-US01-18568-2
: Sequence 2, Application PC/TUS0118568
: GENERAL INFORMATION:
: APPLICANT: Darrow, Andrew L
: APPLICANT: Qi, Jai-shen
: TITLE OF INVENTION: DNA encoding human serine protease D-G
: FILE REFERENCE: ORT-1273
: CURRENT APPLICATION NUMBER: PCT/US01/18568
: PRIORITY FILING DATE: 2001-06-08
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 435
: TYPE: PRT
: ORGANISM: Homo sapiens
: OTHER INFORMATION:
PCT-US01-18568-2
```

```
Query Match 39.5%; Score 521; DB 1; Length 435;
Best Local Similarity 44.6%; Pred. No. 1e-42;
Matches 108; Conservative 36; Mismatches 84; Indels 14; Gaps 7;
```

```
QY 1 RIVGESALPGAMPQVSLHVNHYVCGSITTPETVTAHCHVEKPLNNPMHTAFAGI 60
Db 202 RIVGGEASVDSMPQVSIQYDKOHVCGSITLDPHWVLTAAHCFERRRNVLSRMKVFAGS 260
QY 61 LR-QSFMEYAGYQVEKYVISHPNYDSKTKNNIDIALMKLOKPLTFNDLVPCVLPNG 116
```

```
Db 261 DKIGSF-----PSLAVAKIIIEFNPMY---PKNDIALMKLOPILTFSGTVRPICLPFPD 313
QY 117 MMLQPEQLCWMISGMGNTTEK-GKTSFVLAARVLLIETORCNSRYVYDNLITPAMICAGF 175
Db 314 EELTPATPLMIIIGMGFTKONGKMSDILLQASVQVIDSTRCNADRAYOGGEVTEKMCAGI 373
QY 176 LOGNVDSGCGDSGGPLVTSKNNIMMLIGDTSWMSGCAKAYRPGVYGNVAVFTDVIYRQMR 235
Db 374 PEGGYDTGCGDSGGPLMYSDQ-WHVYGVISMGYCGGPGSTPGVYTKVSATILNMIYNVWK 432
QY 236 AD 237
Db 433 AE 434
```

Search completed: September 26, 2001, 09:23:23
Job time: 138 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 09:21:05 : Search time 20.35 Seconds
(without alignments)
497.811 Million cell updates/sec

Title: US-09-615-285-2
2717
Sequence: 1 MALNCGSPPAIGPYENHGX.....VYGNVWFETDPIYRQMBADG 492

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	492	US-09-342-749-2	Sequence 2, Appl1
2	1540	56.7	283	US-08-807-151-1	Sequence 1, Appl1
3	676.5	24.9	435	US-09-008-271A-6	Sequence 6, Appl1
4	660	24.3	798	US-08-200-900A-2	Sequence 2, Appl1
5	660	24.3	798	PCT-US94-00616-2	Sequence 2, Appl1
6	558.5	20.6	855	US-09-027-337-2	Sequence 2, Appl1
7	556	20.4	638	US-08-681-151-3	Sequence 3, Appl1
8	553.5	19.9	428	US-08-944-483-63	Sequence 63, Appl1
9	539.5	19.9	416	US-09-000-846-2	Sequence 1, Appl1
10	534	19.7	356	US-08-681-151-1	Sequence 1, Appl1
11	528.5	19.5	258	US-09-027-337-3	Sequence 3, Appl1
12	527.5	19.4	416	US-08-508-448C-25	Sequence 25, Appl1
13	524	19.3	238	US-08-944-483-64	Sequence 64, Appl1
14	523.5	19.3	235	US-08-944-483-67	Sequence 67, Appl1
15	521.5	19.2	235	US-08-807-151-3	Sequence 3, Appl1
16	518.5	19.1	812	US-08-248-629A-1	Sequence 1, Appl1
17	518.5	19.1	812	US-08-451-932-1	Sequence 1, Appl1
18	518.5	19.1	812	US-08-452-260-1	Sequence 1, Appl1
19	518.5	19.1	812	US-08-326-785-1	Sequence 1, Appl1
20	518.5	19.1	812	US-08-612-788-1	Sequence 1, Appl1
21	518.5	19.1	812	US-08-605-598B-1	Sequence 1, Appl1
22	518.5	19.1	812	US-08-429-743-1	Sequence 1, Appl1
23	518.5	19.1	812	US-08-866-735-1	Sequence 1, Appl1
24	518.5	19.1	812	US-09-066-028-1	Sequence 1, Appl1
25	518.5	19.1	812	PCT-US95-05107-1	Sequence 1, Appl1
26	513.5	18.9	232	US-08-508-448C-19	Sequence 19, Appl1
27	511.5	18.8	235	US-08-944-483-65	Sequence 65, Appl1

28	495.5	18.2	250	4	US-08-944-483-68	Sequence 68, Appl1
29	459.5	16.9	314	6	US-09-008-271A-3	Sequence 3, Appl1
30	452	16.6	546	6	5200340-6	Patent No. 5200340
31	450	16.6	270	2	US-08-978-404B-8	Sequence 8, Appl1
32	449	16.5	655	1	US-08-148-910-12	Sequence 12, Appl1
33	449	16.5	655	1	US-08-448-937A-12	Sequence 12, Appl1
34	447	16.5	276	2	US-09-016-366A-15	Sequence 15, Appl1
35	447	16.5	276	2	US-08-978-404B-21	Sequence 21, Appl1
36	446	16.4	230	1	US-08-379-621-2	Sequence 2, Appl1
37	446	16.4	230	1	US-08-147-000B-2	Sequence 2, Appl1
38	446	16.4	230	2	US-08-889-078-2	Sequence 2, Appl1
39	445.5	16.4	791	1	US-08-643-219-1	Sequence 1, Appl1
40	445.5	16.4	791	2	US-09-131-995-1	Sequence 1, Appl1
41	445.5	16.4	791	2	US-08-832-087B-1	Sequence 1, Appl1
42	445.5	16.4	791	3	US-08-851-350-1	Sequence 1, Appl1
43	445.5	16.4	791	4	US-09-132-154-1	Sequence 1, Appl1
44	445.5	16.4	810	1	US-07-854-603-2	Sequence 2, Appl1
45	445.5	16.4	810	1	US-08-147-000B-29	Sequence 29, Appl1

ALIGNMENTS

RESULT 1
US-09-342-749-2
; Sequence 2, Appl1 Application US/09342749
; Patent No. 6166194
GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2316-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; EARLIER FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match 100.0%; Score 2717; DB 4; Length 492;
Best Local Similarity 100.0%; Pred. No. 2.1e-241;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALNCGSPPAIGPYENHGXQPNPFPAPQTVVYEVHAPQVYSPVQVAPRVLTQA 60
|||||
DB 1 MALNCGSPPAIGPYENHGXQPNPFPAPQTVVYEVHAPQVYSPVQVAPRVLTQA 60
|||||
QY 61 SNPVVCTOPSPGPGVCTSTKTKALCITTLTGFTLGAALAAGLLKKPKMCKNSNGIEC 120
|||||
DB 61 SNPVVCTOPSPGPGVCTSTKTKALCITTLTGFTLGAALAAGLLKKPKMCKNSNGIEC 120
|||||
QY 121 DSSGTCINPNCMDGVSHCPGDEENRCVLYGPNFTLOYSSQKRSWHPVCCDDNNENY 180
|||||
DB 121 DSSGTCINPNCMDGVSHCPGDEENRCVLYGPNFTLOYSSQKRSWHPVCCDDNNENY 180
|||||
QY 181 GRACADMDGKKNFYSQGLVDDSGSTSPFKLTSAGNVDIYKKLYHSDACSKAVYSLR 240
|||||
DB 181 GRACADMDGKKNFYSQGLVDDSGSTSPFKLTSAGNVDIYKKLYHSDACSKAVYSLR 240
|||||
QY 241 CACGYNLNSRSRSLVYGESALPGAMPQVSLHYQNVHVCSSITTPETVYAAHCVCK 300
|||||
DB 241 CACGYNLNSRSRSLVYGESALPGAMPQVSLHYQNVHVCSSITTPETVYAAHCVCK 300
|||||
QY 301 PLNNPMMWAFACILROSEFVYAGYQVEKVIHPYDSTKKNNDIALKLKLPPLTFNDL 360
|||||

Db 301 PLNNPWHMTAFAGILRQSFMEYGAGYQVEKVISHPNDYSKTKNNIDIALMKLQKPLTFENDL 360
QY 361 VKPVCPLPDPGMMLOPEQOLCWMISGNGATEEKGKTSSEVLNAAKVLLIETORCNSRYVDNLI 420
Db 361 VKPVCPLPDPGMMLOPEQOLCWMISGNGATEEKGKTSSEVLNAAKVLLIETORCNSRYVDNLI 420
QY 421 TPAMICAGFLQGNVDSGCGDSGGLVTSKNNIMWLIGDTSMWGGCAKAYRPGVGNVWF 480
Db 421 TPAMICAGFLQGNVDSGCGDSGGLVTSKNNIMWLIGDTSMWGGCAKAYRPGVGNVWF 480
QY 481 TDWITRQMRADG 492
Db 481 TDWITRQMRADG 492

RESULT 2

US-08-807-151-1
Sequence 1, Application US/08807151

Patent No. 6043033
GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Lal, Preeti

TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/807,151

FILING DATE: Filed Herewith

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PR-0227 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-845-4166

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 283 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: SCORN0701

CLONE: 556016

US-08-807-151-1

Query Match

Best Local Similarity 56.7%; Score 1540; DB 3; Length 283;

Matches 283; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 210 MKLMTSAGNVDIYKLLYSDACSSKAVVSLRCIACGVNLSSRQSRIVGSGALPGANPW 269

Db 1 MKLMTSAGNVDIYKLLYSDACSSKAVVSLRCIACGVNLSSRQSRIVGSGALPGANPW 60

QY 270 QVSLHVQNVHVGCGSIITPEMIVTAHCVKPLNNPWHMTAFAGILRQSFMEYGAGYQVE 329

Db 61 QVSLHVQNVHVGCGSIITPEMIVTAHCVKPLNNPWHMTAFAGILRQSFMEYGAGYQVE 120
QY 330 KYISHPNYDSKTKNNIDIALMKLQKPLTFENDLVKPYCLPDPGMMLOPEQOLCWMISGNGATEE 389
Db 121 KYISHPNYDSKTKNNIDIALMKLQKPLTFENDLVKPYCLPDPGMMLOPEQOLCWMISGNGATEE 180
QY 390 KGTSEVLNAAKVLLIETORCNSRYVDNLIPTAMICAGFLQGNVDSGCGDSGGLVTSK 449
Db 181 KGTSEVLNAAKVLLIETORCNSRYVDNLIPTAMICAGFLQGNVDSGCGDSGGLVTSK 240
QY 450 NNIMWLIGDTSMWGGCAKAYRPGVGNVWFDTWITRQMRADG 492
Db 241 NNIMWLIGDTSMWGGCAKAYRPGVGNVWFDTWITRQMRADG 283

RESULT 3

US-09-008-271A-6
Sequence 6, Application US/09008271A

Patent No. 6203979
GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Hillman, Jennifer L.

Yue, Henry

Guegler, Karl J.

Corley, Neil C.

Tang, Tom Y.

Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/008,271A

FILING DATE: 16-Jan-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: <Unknown>

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheela

REGISTRATION NUMBER: 41,201

REFERENCE/DOCKET NUMBER: PR-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 435 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLN0713

CLONE: 1337018

US-09-008-271A-6

Query Match

Best Local Similarity 24.9%; Score 676.5; DB 4; Length 435;

Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 133 CDGYSHCGRGSDENRCVRLY--GP-----NFIQYISQSKWHIPVCQDDMNENYTG 181

```

Db 75 CDGELDCPLGEDEHCVKSPFEGPAVAVRLSKRSTLLQVDSATGNWFSACFDNFTAL 134
QY 182 RAACDMGY--KNNFYSSO-----GIVDSSGSTRFEMKLTNSAGNVDIYKKLYSDACS 232
Db 135 ETACQOMKYSKPTFRRAEIGPPDDLDVETENSOELMRNSSG-----PCL 182
QY 233 SKAVYSLICAGVALNNSROSRIYGESALPGAMPQVSLHYOVHVCSSGIITPEMIV 292
Db 183 SGLVSLHCLACGESL---KTPRYVGGEEASVDSMPQVSYQYDQHQCGSILDPHWL 239
QY 293 TAAHCVERPLNNPWHMTAFAGILR-QSFMEYAGYQVEKVI---SHPVYDSKTKNNIAL 348
Db 240 TAAHCFERKH-TDVFEMKVRAGSDKLGSE---PSLAVAKIITIEFPWY---PKNDIAL 291
QY 349 MKLQKPLTFNDLVKPVCLPNEGMLQPBOLCISGMGATEEK-GKTSVYLNAAKYLLET 407
Db 292 MKLQPLTFSGVRICLPFEDEELTPATPLMIIIGMFTKQGGMSDILLQASQVIDS 351
QY 408 ORCNSRYVDNLITPAMICAGFLQGNVDSGQSGPLVTSKNNIMWLIGDTSWGSCAK 467
Db 352 TRCNADDAVQGEVETKMKACAGIPEGVDTCQDSSGPLMTQSDQ-WHYVGIYSWGCGG 410
QY 468 AYREGVGNVMTDWTYRQMRAD 491
Db 411 PSTRGVYTKVASAYLWMIYNVWKA 434

RESULT 4
US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 5655566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weinerl, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-200-900A-2

Query Match 24.3%; Score 660; DB 1; Length 798;
Best Local Similarity 33.9%; Pred. No. 3.9e-52;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
QY 65 VCTOPKSPSGIVCTSKTKKALCITLTL-----GTFLVGAALAGLMLKFMKSGCSNSG 117

```

```

Db 374 VYTGPGVNDVSTNNRMTVLFTDNMLAKQGFKANFTTGYLG-----IPECKEDN 426
QY 118 IECDSSTGTCINSNMCDGVSHCPGEGEDENRCVRLYG-----PNFILOYSSQKSMH 169
Db 427 FQC-KDGEICPLVNLCDGPPHCKDGSDEAHCVRLNGLTSSGLVQFRIQ-----SIWH 479
QY 170 PVCODDMWENVRACRDMGYKNNFYSSOGIYVDSGSTRFEMKLTNSAGNVDIYKKLYSD 229
Db 480 VACAENWTTQISDDVQCQLGLGTG--NSSVPFTSTGGCPYVNLNTPNCSLI---LTSQ 534
QY 230 ACSSRAVSLRC-IACGYNLNSROS-RIYGESALPGAMPQVSLHYOVHVCSSGIIT 286
Db 535 QCLEDSLILILQCNYSKCGKLVTOEVSPTIVGSDSRSBAMWVVALFDDQVCGASLY 594
QY 287 TPBWIVTAAHCVERPLNNPWHMTAFAGILRQSFMEYAGYQVE-----KVISHPNYDSKT 341
Db 595 SRDWLVSAHCVYGRNMEPSKWKAVLGLHMASNL---TSPQIETRLIDQIVINPYNKR 651
QY 342 KNDIALMKLOKPLTFNDLVKPVCLPNEGMLQPBOLCISGMGATEEKKTSEVYLNAAK 401
Db 652 KNDIAMMHLKAVNTDITDIPCLPEENQVFPGRICSIAGMGLIYQGSTADYVLEAD 711
QY 402 VLIETORCNSRYVDNLITPAMICAGFLQGNVDSGQSGPLVTSKNNIMWLIGDTSW 461
Db 712 VPLLENKQOQOMPEYN-ITENMVCAGYEAGVDSGQSGPLMCGENRMLLAGVTSF 770
QY 462 GSGCAKATRPQYGVNMTDWTMI 484
Db 771 GYQCALPNRPGYARVPRTETMI 793

RESULT 5
PCT-US94-00616-2
; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-00616-2

Query Match 24.3%; Score 660; DB 5; Length 798;
Best Local Similarity 33.9%; Pred. No. 3.9e-52;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;
QY 65 VCTOPKSPSGIVCTSKTKKALCITLTL-----GTFLVGAALAGLMLKFMKSGCSNSG 117
Db 374 VYTGPGVNDVSTNNRMTVLFTDNMLAKQGFKANFTTGYLG-----IPECKEDN 426
QY 118 IECDSSTGTCINSNMCDGVSHCPGEGEDENRCVRLYG-----PNFILOYSSQKSMH 169
Db 427 FQC-KDGEICPLVNLCDGPPHCKDGSDEAHCVRLNGLTSSGLVQFRIQ-----SIWH 479
QY 170 PVCODDMWENVRACRDMGYKNNFYSSOGIYVDSGSTRFEMKLTNSAGNVDIYKKLYSD 229
Db 480 VACAENWTTQISDDVQCQLGLGTG--NSSVPFTSTGGCPYVNLNTPNCSLI---LTSQ 534

```


Query Match	19.9%	Score 539.5	DB 2	Length 416
Best Local Similarity	30.6%	Pred. No. 1,8e+41		
Matches 132	Conservative	64	Mismatches 179	Indels 57
				Gaps 9
OY	77	CTSKTKKALCTLTCTGTYGACALAGLLMKRDKGSKCSNSGIEDCSGTCINPMSWDCGV	136	
Db	11	CCSRKRVALLVGTLT-LTFTG-----IGASAIYAT	40	
OY	137	SHCPGEGEDENRCVRLTGPENFLQVYSQSRKSWHPYCOQDDMNENYGRACRDMGYKNFYS	196	

```

Db 41 ILGSDQEPYOVQVLSRSLAVFDKTEGTWRLLCSSRSNARVAGLCCCEMGFLRALAH 100
QY 197 SGGIVDDSGSTFEMKLNLSA-----GNVDYKKLYHSDA---CSSKAVSLRACIACGVN 247
Db 101 SELDVRRTAGAN-----GTSGFECVDEGGLRLAQLRLDVIYVDCDCRGRLTITCODCG-- 153
QY 248 LNSSRQSIIVGGSALPGAMPQVSLHVNQVHVCSSITTPEMVITAAHCVEKPLNNPMH 307
Db 154 RRKLPLVDIVIGQDSSLCRMPQVSLRYDGTILCGSLSGMVLTAHCPEERRRVLSR 213
QY 308 WTAFAGLRQSMFYGAGYQVEKYVSHPNY-----DSKTKNDIALKLOKPLTFNDLY 361
Db 214 WRVFGAAVARTSP-HAVDLGVQAVIYHGGYLPFRDPTIDENSNDIALVHLSSSLPLEYI 272
QY 362 KPVCLPFGMMLQPEQLCWMISGMGATEEKGKTSYEVNNAKVLLETORCNRSRYVDNLIT 421
Db 273 QPVCLPAGQALVYDGKCTVYGMWNTQFYGGQAMVLOEARVPIISNENCNSDFEIGNQIK 332
QY 422 PAMICAGFLQGVNDSQDSCGSPLY----TSKNINIMWLIQDTSWGSCKAKAVPGVYGVN 477
Db 333 PKMCAGYPEGGIDACQDQSGPFCEDSISGTSNMRICGIYVSGCALARKPGVYTKV 392
QY 478 MVTMDIYRQMR 489
Db 393 TDFRMTKALK 404

```

```

RESULT 10
US-08-681-151-1
; Sequence 1, Application US/08681151
; Patent No. 5869637
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Braxton, Scott Michael
; APPLICANT: Goll, Surya
; TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,151
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0074US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

```

```

; IMMEDIATE SOURCE:
; LIBRARY: HEARNOT01
; CLONE: 307474
; US-08-681-151-1

```

```

Query Match 19.7%; Score 534; DB 2; Length 356;
Best Local Similarity 35.4%; Pred. No. 4.6e-41;
Matches 120; Conservative 53; Mismatches 132; Indels 34; Gaps 11;

```

```

QY 169 HPVCCDDNNENYGRAACDMGKKNFYSSQGIYDDSGSTSEFKLNT---SAGNDVYKYL 225
Db 25 HHVCDGMOEILISQLACOMGIGEPSYTKL-IQDEKPERKLTLLSNESLNGTTLLELL 83
QY 226 YHSDACSSKAVVSLRCLIA--CGVNLNSSRQSIIVGGSALPGAMPQVSLHVO--NVHVC 282
Db 84 VNGQCESRSKISLTCQDQGXRPAAHMKKRLIGGRISRGPRMWSQSLSEPSGHICG 143
QY 283 GSITPEWIVTRAHCV-----KPLNPMHMTAFAGILRQSTMEYGAGYQVEKY 331
Db 144 CVLAKKMWLVTAHCFEGRENAAYKKVVLGINNLDHPVSF--MOTRF-----VKTI 192
QY 332 ISHPYDSKTKNDIALKLOKPLTFNDLYKPVCLPFGMMLQPEQLCWMISGMGATEEKG 391
Db 193 ILHPYTSRAVVDYDITSEIVLSEDSITETGYRVCPLPNEQWLEPVTYCYITGMG--HMGN 250
QY 392 KTSEVLNNAKVLLETORCNRSRYVDNLITPAMICAGFLQGVNDSQDSCGSPLYTSK-N 450
Db 251 KMPFKLOGEYRIISLEHCGS-YFDMKTIITRMICAGYESGTVSCMDSGSPLYCEKRG 309
QY 451 NIMWLIQDTSWGSCK-AKAYRPGVYGVNMTDWTYRQM 488
Db 310 GRWTLFGLTWSGVCFSKVLPGVYSNVSYFEWIKROI 348

```

```

RESULT 11
US-09-027-337-3
; Sequence 3, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: RADG-15: An Extracellular Serine Protease Overexpressed in
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 3
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)
; OTHER INFORMATION: homologous to similar domain in RADG-15
US-09-027-337-3

```

```

Query Match 19.5%; Score 528.5; DB 2; Length 256;
Best Local Similarity 41.2%; Pred. No. 9.2e-41;
Matches 101; Conservative 41; Mismatches 92; Indels 11; Gaps 3;
QY 255 RIVGESALPGAMPQVSLHVNQVHVCSSITTPEMVITAAHCVEKPLNNPMHTAFAGI 314
Db 1 RIVGGRQTSLSGRMPQVSLRYDGAHLGCSLLSGDWVLTAAHCFERRRVLSRWRVFGA 60
QY 315 LROSMFYGAGYQVEKYVSHPNY-----DSKTKNDIALKLOKPLTFNDLYKPVCLPN 368
Db 61 VAOASP-HGDLGVQAVVYHGGYLPFRDPNSENSENDAIALVHLSSPLPLEYIQVCLPA 119
QY 369 PGMMMLQPEQLCWMISGMGATEEKGKTSYEVNNAKVLLETORCNRSRYVDNLITPAMICAG 428
Db 120 AGQALVYDGKICTVYGMWNTQFYGGQAMVLOEARVPIISNDVCGNADRFIGNQIKRMFCAG 179

```


Db 1 IVGSDSREGAMPWVVALYFDDQVCASLVSRDMLVSAHCVYGRNMEPSKKAVLGLH 60
QY 316 ROSEMFYAGAYOVE-----KVISHPRYDSKTKNNNDIALMKIQKPLFENDLVKPVCLPNPG 370
Db 61 MASNL---TSPQIETRLIDQIVINPHYNNRKRKNDIAMHLEKKNVYTDYIOPICLPEEN 117
QY 371 MMLQPEOLCWMISGWGATEEKGKTSEVLNAAKVLLETFQRNSRYVYDNLITPAMICAGFL 430
Db 118 QYFPPGRICSIAGWGALLIQGSTADVLQEADVPILSNKCCQOQMPPEYN-ITENMVCAGYE 176
QY 431 QGNVDSGCGDSGGLPYTSKNNIMWLIIGTWSGSGCAKAYRPGVYGNVWVFTDNI 484
Db 177 AGGVDSGCGDSGGLPYTSKNNIMWLIIGTWSGSGCAKAYRPGVYGNVWVFTDNI 230

Search completed: September 26, 2001, 09:22:08
Job time: 63 sec

THIS PAGE BLANK (USPTO)